

EXHIBIT 5

Modern Industrial Organization

Fourth Edition

Dennis W. Carlton

University of Chicago

Jeffrey M. Perloff

University of California, Berkeley



Boston San Francisco New York
London Toronto Sydney Tokyo Singapore Madrid
Mexico City Munich Paris Cape Town Hong Kong Montreal

To Janie and Jackie

Editor in Chief: Denise Clinton
Acquisitions Editor: Adrienne D'Ambrosio
Director of Development: Sylvia Mallory
Managing Editor: Jim Rigney
Senior Production Supervisor: Nancy Fenton
Senior Media Producer: Melissa Honig
Marketing Manager: Deborah Meredith
Design Supervisor: Regina Kolenda
Interior Designer: Leslie Haimes
Cover Designer: Leslie Haimes
Illustrator: Jim McLaughlin
Senior Prepress Supervisor: Caroline Fell
Senior Manufacturing Buyer: Hugh Crawford
Compositor: Nesbitt Graphics, Inc.
Cover Image: Stella, Frank (b. 1936) © Copyright ARS, NY. Agbatana II. 1968. Polymer and fluorescent polymer paint on canvas. 10' × 15'. © Copyright ARS, NY. Musee d'Art et d'Industrie, Saint-Etienne. Photo credit: Giraudon

Many of the designations used by manufacturers and sellers to distinguish their products are claimed as trademarks. Where those designations appear in this book and Addison-Wesley is aware of a trademark claim, the designations have been printed in initial caps or all caps.

Library of Congress Cataloging-in-Publication Data

Carlton, Dennis W.

Modern industrial organization / Dennis W. Carlton, Jeffrey M. Perloff.—
4th ed.

p. cm

Includes bibliographical references and index.

ISBN 0-321-18023-2

1. Industrial organization (Economic theory) I. Perloff, Jeffrey M. II.

Title

HD2326.C376 2005

338.6—dc22

2004007014

Copyright © 2005 by Dennis W. Carlton and Jeffrey M. Perloff

All rights reserved. No part of this publication may be reproduced, stored in a database or retrieval system, distributed, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the publisher. Printed in the United States of America.

1 2 3 4 5 6 7 8 9 10—CRW—08 07 06 05 04

Brief Contents

PART 1 *Introduction and Theory*

- CHAPTER 1 Overview
- CHAPTER 2 The Firm and Costs

PART 2 *Market Structures*

- CHAPTER 3 Competition
- CHAPTER 4 Monopolies, Monopsonies, and Dominant Firms
- CHAPTER 5 Cartels
- CHAPTER 6 Oligopoly
- CHAPTER 7 Product Differentiation and Monopolistic Competition
- CHAPTER 8 Industry Structure and Performance

PART 3 *Business Practices: Strategies and Conduct*

- CHAPTER 9 Price Discrimination
- CHAPTER 10 Advanced Topics in Pricing
- CHAPTER 11 Strategic Behavior
- CHAPTER 12 Vertical Integration and Vertical Restrictions

PART 4 *Information, Advertising, and Disclosure*

- CHAPTER 13 Information
- CHAPTER 14 Advertising and Disclosure

PART 5 *Dynamic Models and Market Clearing*

- CHAPTER 15 Decision Making over Time: Durability
- CHAPTER 16 Patents and Technological Change
- CHAPTER 17 How Markets Clear: Theory and Facts

PART 6 *Government Policies and Their Effects*

- CHAPTER 18 International Trade
- CHAPTER 19 Antitrust Laws and Policy
- CHAPTER 20 Regulation and Deregulation

Contents

PART 1 *Introduction and Theory* 1

CHAPTER 1	Overview	2	Economies of Scale	36
	Models	2	Reasons for Economies of Scale	36
	Price Theory	3	Total Costs Determine Scale Economies	38
	Transaction Costs	5	A Measure of Scale Economies	39
	Game Theory	6	Empirical Studies of Cost Curves	40
	Contestable Markets	6	Economies of Scale in Total Manufacturing Costs	40
	Organization	6	Survivorship Studies	42
	Basic Theory	7	Cost Concepts for Multiproduct Firms	43
	Market Structures	7	Adaptation of Traditional Cost Concepts for a Multiproduct Firm	44
	Business Practices: Strategies and Conduct	9	Economies of Scope	44
	Information, Advertising, and Disclosure	9	Economies of Scale and Economies of Scope	45
	Dynamic Models and Market Clearing	10	Specialization in Manufacturing	45
	Government Policies and Their Effects	10	An Example of an Industry with Economies of Scope	46
CHAPTER 2	The Firm and Costs	11	Summary	47
	The Firm	12	Problems	48
	The Objective of a Firm	12	Suggested Readings	49
	Ownership and Control	13	APPENDIX 2A Cost Concepts for a Multiproduct Firm	50
	Mergers and Acquisitions	20	EXAMPLE 2.1 Value of Limited Liability	15
	Reasons for Mergers and Acquisitions	20	EXAMPLE 2.2 Conflicts of Interest Between Managers and Shareholders	18
	Merger Activity in the United States	23	EXAMPLE 2.3 Specialization of Labor	37
	Merger Activities in Other Countries	25	EXAMPLE 2.4 Indiana Libraries	41
	Empirical Evidence on the Efficiency and Profitability of Mergers	26	EXAMPLE 2.5 The Baking Industry	46
	Cost Concepts	29	EXAMPLE 2.6 Electricity Minimum Efficient Scale and Scope	47
	Types of Costs	29		
	Cost Concepts	33		

PART 2 *Market Structures* 55

CHAPTER 3	Competition	56	CHAPTER 4	Monopolies, Monopsonies, and Dominant Firms	88
Perfect Competition		57	Monopoly Behavior		89
Assumptions		57	Profit Maximization		89
The Behavior of a Single Firm		58	Market and Monopoly Power		93
The Competitive Market		61	The Incentive for Efficient Operation		94
Elasticities and the Residual Demand Curve		65	Monopoly Behavior over Time		94
Elasticities of Demand and Supply		65	The Costs and Benefits of Monopoly		95
The Residual Demand Curve of Price Takers		66	The Deadweight Loss of Monopoly		95
Efficiency and Welfare		69	Rent-Seeking Behavior		96
Efficiency		70	Monopoly Profits and Deadweight Loss Vary with the Elasticity of Demand		97
Welfare		70	The Benefits of Monopoly		99
Entry and Exit		73	Creating and Maintaining a Monopoly		99
Restrictions on Entry		73	Knowledge Advantage		100
Competition with Few Firms—Contestability		76	Government-Created Monopolies		102
Definition of Barriers to Entry		76	Natural Monopoly		104
Identifying Barriers to Entry		79	Profits and Monopoly		105
The Size of Entry Barriers by Industry		80	Is Any Firm That Earns a Positive Profit a Monopoly?		105
Externalities		82	Does a Monopoly Always Earn a Positive Profit?		106
Limitations of Perfect Competition		84	Are Monopoly Mergers to Eliminate Short- Run Losses Desirable?		106
The Many Meanings of Competition		85	Monopsony		107
Summary		86	Dominant Firm with a Competitive Fringe		110
Problems		86	Why Some Firms Are Dominant		111
EXAMPLE 3.1 Are Farmers Price Takers?		69	The No-Entry Model		112
EXAMPLE 3.2 Restrictions on Entry Across Countries		74	The Dominant Firm—Competitive Fringe Equilibrium		115
EXAMPLE 3.3 FTC Opposes Internet Bans That Harm Competition		75	A Model with Free, Instantaneous Entry		116
EXAMPLE 3.4 Increasing Congestion		83			

Summary	119	EXAMPLE 5.4 Relieving the Headache of Running a Cartel	140
Problems	120	EXAMPLE 5.5 Vitamins Cartel	142
Suggested Readings	121	EXAMPLE 5.6 How Consumers Were Railroaded	145
EXAMPLE 4.1 Monopoly Newspaper Ad Prices	93	EXAMPLE 5.7 The Social Costs of Cartelization	150
EXAMPLE 4.2 Monopolizing by Merging	100	EXAMPLE 5.8 Prosecuting Global Cartels	152
EXAMPLE 4.3 Controlling a Key Ingredient	101		
EXAMPLE 4.4 Preventing Imitation—Cat Got Your Tongue?	101	CHAPTER 6 Oligopoly	157
EXAMPLE 4.5 Protecting a Monopoly	103	Game Theory	159
EXAMPLE 4.6 EU Allows Merger to Eliminate Losses	106	Single-Period Oligopoly Models	160
EXAMPLE 4.7 Priest Monopsony	109	Nash Equilibrium	161
EXAMPLE 4.8 Price Umbrella	111	The Cournot Model	161
EXAMPLE 4.9 China Tobacco Monopoly to Become a Dominant Firm	117	The Bertrand Model	171
		The Stackelberg Leader-Follower Model	176
		A Comparison of the Major Oligopoly Models	180
CHAPTER 5 Cartels	122	Multiperiod Games	180
Why Cartels Form	123	Single-Period Prisoners' Dilemma Game	181
Creating and Enforcing the Cartel	125	Infinitely Repeated Prisoners' Dilemma Game	183
Factors That Facilitate the Formation of Cartels	127	Types of Equilibria in Multiperiod Games	186
Enforcing a Cartel Agreement	136	Experimental Evidence on Oligopoly Models	189
Cartels and Price Wars	144	Summary	191
Consumers Gain as Cartels Fail	147	Problems	192
Price-Fixing Laws	150	Suggested Readings	193
Summary	152	APPENDIX 6A A Mathematical Derivation of Cournot and Stackelberg Equilibria	194
Problems	154	APPENDIX 6B Mixed Strategies	198
Suggested Readings	154	EXAMPLE 6.1 Do Birds of a Feather Cournot- Flock Together?	166
APPENDIX 5A The Effects of Cartel Size	155	EXAMPLE 6.2 Oligopoly Welfare Losses	169
EXAMPLE 5.1 An Electrifying Conspiracy	128	EXAMPLE 6.3 Mergers in a Cournot Economy	171
EXAMPLE 5.2 The Viability of Commodity Cartels	132	EXAMPLE 6.4 Roller Coaster Gasoline Pricing	177
EXAMPLE 5.3 Concrete Example of Government Aided Collusion	138	EXAMPLE 6.5 Copying Pricing	184
		EXAMPLE 6.6 Car Wars	185

CHAPTER 7 Product Differentiation and Monopolistic Competition	200	EXAMPLE 7.4 A Serial Problem	229
		EXAMPLE 7.5 Combining Beers	232
Differentiated Products	202	EXAMPLE 7.6 Value of Minivans	233
The Effect of Differentiation on a Firm's Demand Curve	203	CHAPTER 8 Industry Structure and Performance	244
Preferences for Characteristics of Products	205		
The Representative Consumer Model	206	Theories of Price Markups and Profits	245
A Representative Consumer Model with Undifferentiated Products	206	Structure-Conduct-Performance	246
A Representative Consumer Model with Differentiated Products	214	Measures of Market Performance	246
Conclusions About Representative Consumer Models	219	Rates of Return	247
Location Models	220	Price-Cost Margins	254
Hotelling's Location Model	221	Measures of Market Structure	255
Salop's Circle Model	223	The Relationship of Structure to Performance	259
Hybrid Models	230	Modern Structure-Conduct-Performance Analysis	268
Estimation of Differentiated Goods Models	231	Theory	268
Summary	233	Empirical Research	272
Problems	234	Modern Approaches to Measuring Performance	274
Suggested Readings	234	Static Studies	275
APPENDIX 7A Welfare in a Monopolistic Competition Model with Homogeneous Products	235	Summary	281
APPENDIX 7B Welfare in a Monopolistic Competition Model with Differentiated Products	240	Problems	282
EXAMPLE 7.1 All Water Is Not the Same	204	APPENDIX 8A Relationship Between the Herfindahl-Hirschman Index (HHI) and the Price-Cost Margin	283
EXAMPLE 7.2 Entry Lowers Prices	211	APPENDIX 8B Identifying Market Power	284
EXAMPLE 7.3 The Jeans Market	216	EXAMPLE 8.1 Supermarkets and Concentration	273
		EXAMPLE 8.2 How Sweet It Is	278

PART 3 *Business Practices: Strategies and Conduct* 289

CHAPTER 9 Price Discrimination	290	Conditions for Price Discrimination	294
		Resales	294
Nonuniform Pricing	291	Types of Price Discrimination	296
Incentive and Conditions for Price Discrimination	293	Perfect Price Discrimination	299
Profit Motive for Price Discrimination	293	Each Consumer Buys More Than One Unit	300

Different Prices to Different Groups	301	Summary	338
Other Methods of Third-Degree Price Discrimination	305	Problems	340
Welfare Effects of Price Discrimination	306	APPENDIX 10A The Optimal Two-Part Tariff	341
Summary	308	APPENDIX 10B Nonlinear Pricing with an Example	344
Problems	309	EXAMPLE 10.1 Football Tariffs	314
Suggested Readings	309	EXAMPLE 10.2 You Auto Save from Tie-in Sales	320
APPENDIX 9A An Example of Price Discrimination: Agricultural Marketing Orders	310	EXAMPLE 10.3 Stuck Holding the Bag	320
EXAMPLE 9.1 Coupons	292	EXAMPLE 10.4 Tied to TV	325
EXAMPLE 9.2 Thank You, Doctor	295	EXAMPLE 10.5 Not Too Suite—Mixed Bundling	329
EXAMPLE 9.3 Halting Drug Resale from Canada	297	EXAMPLE 10.6 Price Discriminating on eBay	339
EXAMPLE 9.4 Vertical Integration as a Means of Price Discrimination: Alcoa Shows Its True Metal	298	CHAPTER 11 Strategic Behavior	350
EXAMPLE 9.5 A Discriminating Labor Union	302	Strategic Behavior Defined	350
EXAMPLE 9.6 Does Competition Always Lower Price?	308	Noncooperative Strategic Behavior	351
CHAPTER 10 Advanced Topics in Pricing	313	Predatory Pricing	352
Nonlinear Pricing	313	Limit Pricing	360
A Single Two-Part Tariff	314	Investments to Lower Production Costs	367
Two Two-Part Tariffs	316	Raising Rivals' Costs	371
Tie-in Sales	319	Welfare Implications and the Role of the Courts	378
General Justifications for Tie-in Sales	319	Cooperative Strategic Behavior	379
Tie-in Sales as a Method of Price Discrimination	321	Practices That Facilitate Collusion	379
Package Tie-in Sales of Independent Products	322	Cooperative Strategic Behavior and the Role of the Courts	386
Interrelated Demands	330	Summary	386
Quality Choice	336	Problems	387
Other Methods of Nonlinear Pricing	336	Suggested Readings	388
Minimum Quantities and Quantity Discounts	336	APPENDIX 11A: The Strategic Use of Tie-in Sales and Product Compatibility to Create or Maintain Market Power with Applications to Networks	389
Selection of Price Schedules	337	EXAMPLE 11.1 Supreme Court Says Alleged Predation Must Be Credible	354
Premium for Priority	337	EXAMPLE 11.2 Evidence of Predatory Pricing in Tobacco	358
Auctions	338		

EXAMPLE 11.3 The Shrinking Share of Dominant Firms	365	Vertical Restrictions Used to Solve Problems in Distribution	415
EXAMPLE 11.4 And Only a Smile Remained	366	The Effects of Vertical Restrictions	425
EXAMPLE 11.5 Strategic Behavior and Rapid Technological Change: The Microsoft Case	374	Banning Vertical Restrictions	430
EXAMPLE 11.6 Value of Preventing Entry	377	Franchising	431
EXAMPLE 11.7 The FTC versus Ethyl et al.	382	Empirical Evidence	433
EXAMPLE 11.8 Information Exchanges: The Hardwood Case	383	Evidence on Vertical Integration	433
		Evidence on Vertical Restrictions	435
		Summary	437
		Problems	438
		Suggested Readings	438
CHAPTER 12 Vertical Integration and Vertical Restrictions	395	EXAMPLE 12.1 Outsourcing	397
The Reasons for and Against Vertical Integration	396	EXAMPLE 12.2 Preventing Holdups	398
Integration to Lower Transaction Costs	400	EXAMPLE 12.3 Own Your Own Steel Mill	404
Integration to Assure Supply	403	EXAMPLE 12.4 Double Markup	419
Integration to Eliminate Externalities	404	EXAMPLE 12.5 Blockbuster's Solution to the Double Marginalization Problem	420
Integration to Avoid Government Intervention	404	EXAMPLE 12.6 Free Riding on the Web	422
Integration to Increase Monopoly Profits	405	EXAMPLE 12.7 Brewing Trouble: Restricting Vertical Integration in Alcoholic Beverage Industries	426
Integration to Eliminate Market Power	412		
The Life Cycle of a Firm	412		
Vertical Restrictions	414		

PART 4 *Information, Advertising, and Disclosure* 439

CHAPTER 13 Information	440	Providing Consumer Information	
Why Information Is Limited	441	Lowers Price	462
Limited Information About Quality	443	How Information Lowers Prices	463
The Market for "Lemons"	443	An Example: Grocery Store Information Programs	465
Solving the Problem: Equal Information	446	Summary	467
Evidence on Lemons Markets	450	Problems	468
Limited Information About Price	452	Suggested Readings	468
The Tourist-Trap Model	453	APPENDIX 13A Market Shares in the Tourists-and-Natives Model	469
★The Tourists-and-Natives Model	457		

EXAMPLE 13.1 Genetically Modified Organisms: Do Consumers Not Care or Not Read?	442	Advertising to Solve the Lemons Problem	481
EXAMPLE 13.2 Understanding Consumer Information	444	When Advertising Is Excessive	483
EXAMPLE 13.3 Counterfeit Halal Meat	447	False Advertising	489
EXAMPLE 13.4 Certifying Thoroughbreds	451	Limits to Lying	489
EXAMPLE 13.5 Price Dispersion and Search Costs in the Talmud	453	Antifraud Laws	491
EXAMPLE 13.6 Price Dispersion	458	Disclosure Laws	492
EXAMPLE 13.7 Tourist Cameras	464	Summary	494
		Problems	494
		Suggested Readings	495
CHAPTER 14 Advertising and Disclosure	471	APPENDIX 14A Profit-Maximizing Advertising	496
Information and Advertising	474	EXAMPLE 14.1 Branding and Labeling	475
Promotions	474	EXAMPLE 14.2 Celebrity Endorsements	478
“Search” Versus “Experience” Goods	475	EXAMPLE 14.3 Milk Advertising	480
Informational Versus Persuasive Advertising	476	EXAMPLE 14.4 Social Gain from Price Advertising	482
Profit-Maximizing Advertising	477	EXAMPLE 14.5 Welfare Effects of Restricting Alcohol Ads	487
Effects of Advertising on Welfare	480	EXAMPLE 14.6 Let All Who Are Hungry Come and Eat	493
Price Advertising Increases Welfare	481		

PART 5 *Dynamic Models and Market Clearing* 497

CHAPTER 15 Decision Making over Time: Durability	498	Suggested Readings	521
How Long Should a Durable Good Last?	498	APPENDIX 15A Multiperiod Durable Goods Monopoly	522
Competitive Firm’s Choice of Durability	499	EXAMPLE 15.1 United Shoe	504
The Monopoly’s Choice of Durability	500	EXAMPLE 15.2 The Importance of Used Goods	508
Costly Installation and Maintenance	503	EXAMPLE 15.3 The Alcoa Case: Secondhand Economics	510
Renting Versus Selling by a Monopoly	505	EXAMPLE 15.4 Leasing Under Adverse Selection	512
Resale Market	505	EXAMPLE 15.5 Sales Versus Rentals	519
★ Consumers’ Expectations Constrain the Monopoly	507	EXAMPLE 15.6 Lowering the Resale Price of Used Textbooks	520
Summary	520		
Problems	521		

CHAPTER 16 Patents and Technological Change	525	EXAMPLE 16.6 Mickey Mouse Legislation	552
		EXAMPLE 16.7 European Patents	555
Patents, Copyrights, and Trademarks	526	EXAMPLE 16.8 Patent Thicket	565
Patents	526		
Copyrights	528	CHAPTER 17 How Markets Clear: Theory and Facts	569
Trademarks	529		
Distinctions Between Patents, Copyrights, and Trademarks	530	How Markets Clear: Three Simple Theories	570
Incentives for Inventions Are Needed	530	Competition	570
Imitation Discourages Research	532	Oligopoly Models	570
Patents Encourage Research	532	Monopoly	570
Patents Encourage Disclosure	536	Empirical Evidence on the Role of Price in Allocating Goods	571
★ Patents, Prizes, Research Contracts, and Joint Ventures	539	The Rigidity of Prices	572
Determining the Optimal Number of Firms	541	Movements in Prices and Price-Cost Margins over the Business Cycle	578
No Government Incentives	542	Explaining the Evidence	579
Government-Financed Research	543	Extensions to the Simple Theory: The Introduction of Time	580
Prizes	544	Fixed Costs of Changing Price	583
Relaxing Antitrust Laws: Joint Ventures	547	Implications of an Unchanging Price for Inventories	584
Patents	548	Asymmetric Information and Moral Hazard	586
Government Uncertainty	554	Toward a General Theory of Allocation	587
Patent Holders May Manufacture or License	554	Market Structure Is More Than Concentration	591
Eliminating Patents	558	Produce-to-Order Versus Produce-to-Stock	591
Market Structure	560	Transmission of Shocks in Industries with Fixed Prices	592
Market Structure Without a Patent Race	560	Summary	593
Optimal Timing of Innovations	561	Problems	593
Monopolies in Patent Races	564	EXAMPLE 17.1 Price Rigidity: It's the Real Thing	577
Summary	566	EXAMPLE 17.2 How Much Is That Turkey in the Window?	579
Problems	567	EXAMPLE 17.3 The Cost of Changing Prices	585
Suggested Readings	568	EXAMPLE 17.4 Creating Futures Markets	588
EXAMPLE 16.1 Piracy	534	EXAMPLE 17.5 Oh Say Does That Star-Spangled Banner Yet Fly?	590
EXAMPLE 16.2 Patents Versus Secrets	537		
EXAMPLE 16.3 Monkey See, Monkey Do	538		
EXAMPLE 16.4 Joint Public-Private R&D	544		
EXAMPLE 16.5 Prizes	545		

PART 6 *Government Policies and Their Effects* 595

CHAPTER 18 International Trade	596	The Use of U.S. Antitrust Laws	640
Reasons for Trade Between Countries	596	Private Litigation	642
Comparative Advantage	597	Market Power and the Definition of Markets	642
Intra-Industry Trade in Differentiated Products	598	Market Power	642
Free Riding, International Price Differences, and Gray Markets	598	Market Definition	644
Dumping	603	Cooperation Among Competitors	648
Tariffs, Subsidies, and Quotas	608	Price-Fixing and Output Agreements	649
Competition	609	Not All Agreements Among Competitors Are Illegal	651
Creating and Battling Monopolies	611	Information Exchanges Among Competitors	653
Strategic Trade Policy	618	Oligopoly Behavior	655
Industries with Positive Externalities	622	Mergers	657
Empirical Evidence on Intervention in International Trade	623	Exclusionary Actions and Other Strategic Behavior	661
Summary	627	Competition Between Rivals	662
Problems	627	Competitive Behavior Deemed Undesirable by the Court	662
APPENDIX 18A Derivation of the Optimal Subsidy	629	Vertical Arrangements Between Firms	668
EXAMPLE 18.1 Gray Markets	600	Price Discrimination	674
EXAMPLE 18.2 Timber Wars and Retaliation	612	Price Discrimination Under Robinson-Patman	674
EXAMPLE 18.3 Foreign Doctors	616	Tie-in Sales	675
EXAMPLE 18.4 Being Taken for a Ride: Japanese Cars	618	Effects of Antitrust Laws on the Organization of Unregulated and Regulated Firms	677
EXAMPLE 18.5 Wide-Body Aircraft	625	Summary	679
EXAMPLE 18.6 Stealing from U.S. Customers	626	Problems	680
CHAPTER 19 Antitrust Laws and Policy	631	Suggested Readings	681
The Antitrust Laws and Their Purposes	632	EXAMPLE 19.1 Using the Government to Create Market Power: Misuse of the Orange Book	635
Antitrust Statutes	632	EXAMPLE 19.2 Conflict Between European and U.S. Antitrust Authorities: GE-Honeywell	638
Enforcement	633	EXAMPLE 19.3 The Merger Guidelines	645
Goals of the Antitrust Laws	634		
Who May Sue?	637		
Economic Theory of Damages	639		

EXAMPLE 19.4 Antitrust Laws in Other Countries	650	Summary	734
EXAMPLE 19.5 Colleges and Antitrust: Does Your School Belong to a Cartel?	654	Problems	734
EXAMPLE 19.6 The FTC Plays with Toys 'Я' Us	673	Suggested Readings	735
CHAPTER 20 Regulation and Deregulation	682	EXAMPLE 20.1 Pizza Protection	685
The Objectives of Regulators	686	EXAMPLE 20.2 Cross-Subsidization	688
Market Inefficiencies	686	EXAMPLE 20.3 Legal Monopolies	689
Correcting Market Inefficiencies	687	EXAMPLE 20.4 Public, Monopolistic, and Competitive Refuse Collection	693
Capture Theory and Interest-Group Theory	687	EXAMPLE 20.5 Rent Control	715
Making Monopolies More Competitive	691	EXAMPLE 20.6 Brewing Trouble	717
Government Ownership	692	EXAMPLE 20.7 Deregulating Electricity: California in Shock	726
Privatizing	692	EXAMPLE 20.8 International and U.S. Deregulation in Telecommunications	728
Franchise Bidding	694	EXAMPLE 20.9 European Deregulation of Airlines	731
Price Controls	696	Bibliography	737
★Rate-of-Return Regulation	707	Glossary	780
Quality Effects	712	Answers to Odd-Numbered Problems	787
Making Competitive Industries More Monopolistic	714	Legal Case Index	795
Limiting Entry	716	Author Index	798
Agricultural Regulations: Price Supports and Quantity Controls	718	Subject Index	808
Deregulation	721		
Airlines	724		
Ground Transportation	730		

Monopolies, Monopsonies, and Dominant Firms

Aeroflot Airlines: You Have Made the Right Choice.

—Ad campaign for the only airline in the then Soviet Union

A firm is a **monopoly** if it is the only supplier of a product for which there is no close substitute. A monopoly sets its price without fear that it will be undercut by a rival firm. A monopoly faces a downward-sloping demand curve and sets a price above marginal cost. As a result, less is sold than if the market were competitive (where price equals marginal cost) and society suffers a deadweight loss.

This chapter analyzes a monopoly's behavior and the consequences of that behavior. It also discusses how a monopoly is maintained and asks whether monopoly is always bad. The effects of externalities in a monopolized market are discussed next. The chapter then turns to two related topics. It examines monopsony, which is a monopoly on the buying side of the market. Then it discusses what happens to a monopoly if higher-cost competitive firms enter its market.

The six key questions we answer in this chapter are:

1. How does monopoly compare to competition in terms of prices and welfare?
2. How are monopolies created and maintained?
3. Are there markets in which there are benefits to monopoly?
4. Are all firms that earn profits monopolies, do all monopolies earn profits, and can monopolies earn profits in the long run?
5. How does a monopsony exercise its market power?
6. What happens to a monopoly if smaller, price-taking firms enter its market?

Monopoly Behavior

Because a monopoly faces a downward-sloping market demand curve, it can raise its price above marginal cost. To maximize its profit, it has an incentive to produce its output efficiently. A firm's behavior and government regulations influence the firm's ability to become and remain a monopoly.

Profit Maximization

Price, n. Value, plus a reasonable sum for the wear and tear of conscience in demanding it.
—Ambrose Bierce

Like a competitive firm, a monopoly sets its level of output to maximize its profits. Because the market demand curve is downward sloping, the more the monopoly sells, the lower the price it receives.

The market demand curve constrains the monopoly. In its quest to maximize profit, it can set only price or only quantity—not both. If the monopoly sets quantity, the market price is determined by the market demand curve. If it sets price, the quantity is determined by the market demand curve.

Given the demand curve in Figure 4.1, if the monopoly wants to sell Q_0 units of its product, it charges price p_0 . If it wishes to sell one more unit, it has to lower its price to p_1 .

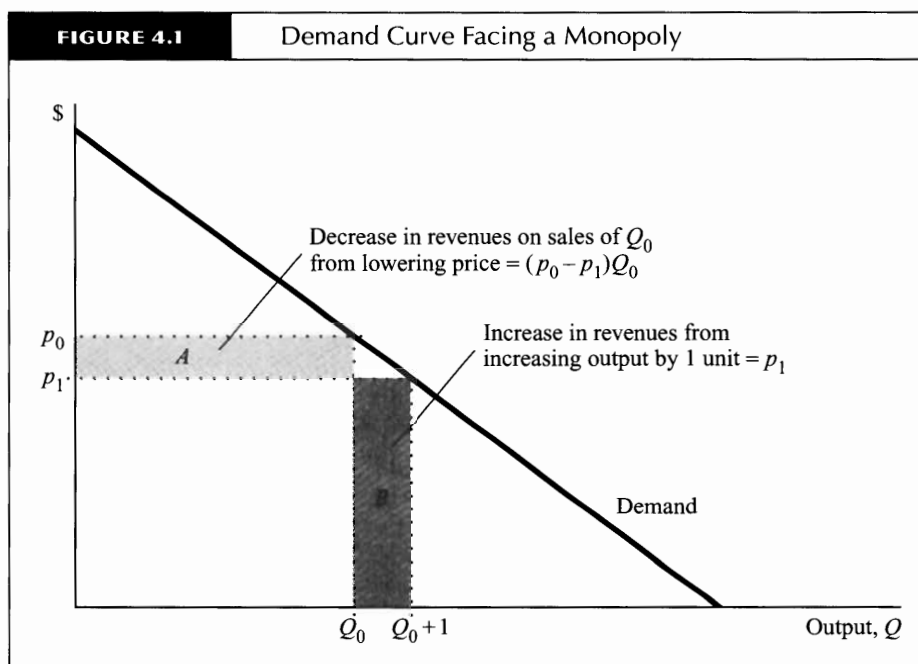
If the monopoly lowers its price to p_1 , its revenues may rise or fall. The monopoly gains revenue on the extra unit it sells at price p_1 , Area B in Figure 4.1. To sell that extra unit, however, it must cut its price from p_0 to p_1 on the original Q_0 units, resulting in a loss of revenues of $(p_0 - p_1)Q_0$, Area A in Figure 4.1.

When discussing a competitive firm's behavior in Chapter 3, we did not have to consider this loss of revenue due to lower price. Because a competitive, price-taking firm faces a horizontal demand curve, the price it receives does not fall if it expands its quantity.

If Area B is larger than Area A in Figure 4.1, then selling the extra unit causes revenues to rise. The extra revenues, $p_1(Q_0 + 1) - p_0Q_0$, that a firm receives when it produces one more unit of the product is called the **marginal revenue**.¹ Hence, the marginal revenue equals area B minus Area A . If the monopoly did not have to lower its price to sell the additional unit, then the increment to revenues from selling an additional unit would simply be the initial price, p_0 . But because the demand curve

¹Marginal revenue, MR , is the change in revenue from selling an additional unit. Total revenue is $p(Q)Q$, where $p(Q)$ is the inverse demand curve (p is a decreasing function of Q). Marginal revenue is equal to $p + Q(\Delta p/\Delta Q)$, where $(\Delta p/\Delta Q)$ is the decline in price necessary to sell the additional unit. Using calculus,

$$MR = \frac{d(p(Q)Q)}{dQ} = p + \frac{dp}{dQ} Q.$$



is downward sloping, the monopoly must lower its price to sell more units. Therefore, the marginal revenue is always less than the price for a monopoly, as Figure 4.2a illustrates.² For a firm in a perfectly competitive market, marginal revenue equals price.

Marginal revenue and total revenue are closely related. When marginal revenue is positive, total revenue increases as output expands, but when marginal revenue is negative, total revenue falls as output expands. As a result, total revenues are maximized (Figure 4.2b) when marginal revenue equals zero (Figure 4.2a).³

A monopoly maximizes its profit rather than its revenue (just as a competitive firm does). Profit is maximized at a smaller quantity than is revenue, as Figure 4.2b illustrates.

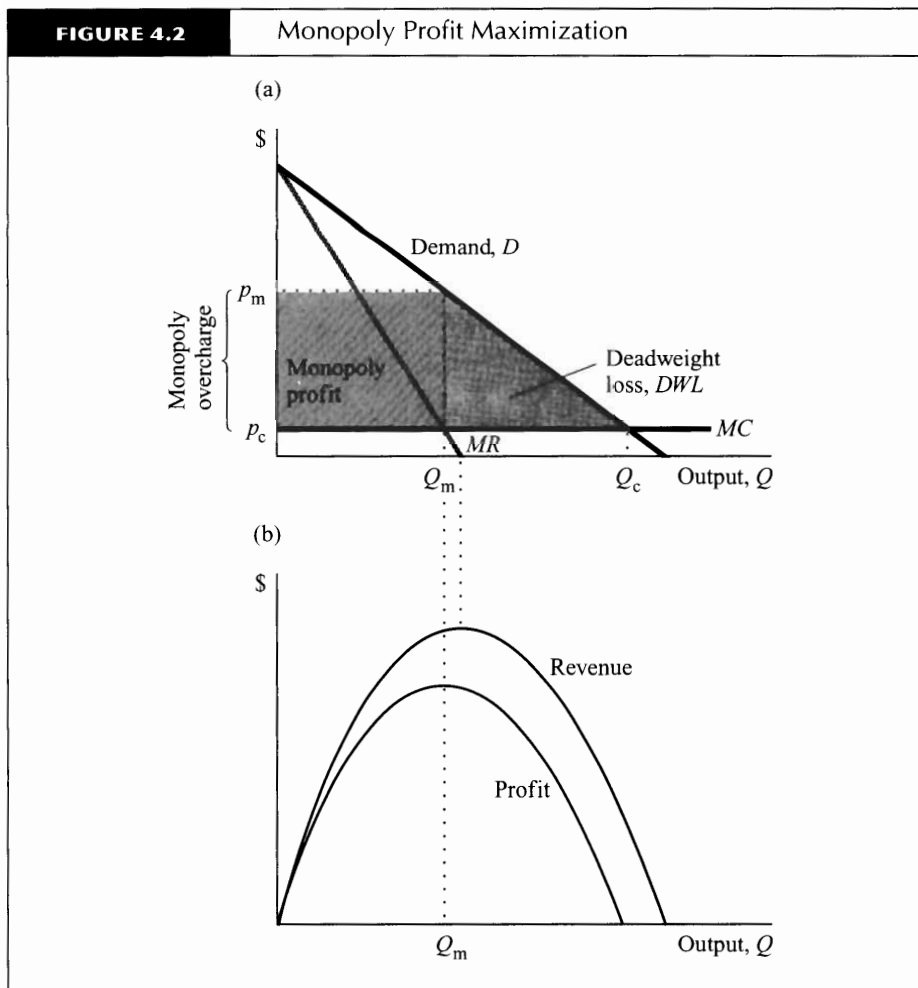
²If a straight-line demand curve, as in Figure 4.2a, hits a horizontal line at Q , the corresponding marginal revenue curve is also a straight line and hits the horizontal line at $Q/2$. To prove this result, let the straight-line demand curve be $p = a - bQ$. Total revenue is $R = pQ = aQ - bQ^2$. The marginal revenue curve is obtained by differentiating R with respect to Q : $MR = a - 2bQ$. The demand curve hits the horizontal axis ($p = 0$) at $Q = a/b$. The marginal revenue curve hits the horizontal axis ($MR = 0$) at $Q = a/(2b)$.

³If the monopoly wants to maximize revenues through its choice of Q ,

$$\max_Q R = p(Q)Q,$$

it sets its marginal revenues equal to zero (first-order condition):

$$MR = p + \frac{dp}{dQ}Q = 0.$$



A monopoly maximizes its profit when the extra revenue from selling one more unit just equals the extra cost of producing that last unit of output. That is, profit is maximized where *marginal revenue equals marginal cost*.⁴

$$MR = MC. \quad (4.1)$$

⁴If the monopoly wants to maximize profits through its choice of Q

$$\max_Q \pi = p(Q)Q - C(Q),$$

it sets its marginal profits equal to zero (first-order condition):

$$\frac{d\pi}{dQ} = MR - MC = \left(p + \frac{dp}{dQ} Q \right) - \frac{dC}{dQ} = 0. \quad (\text{continues})$$

Figure 4.2a illustrates this profit-maximizing relationship. The profit-maximizing monopoly output, Q_m , is smaller than the competitive output, Q_c , determined by the intersection of the demand curve with the marginal cost curve (which we assume would be the supply curve if the market were competitive) at price p_c . The monopoly does *not* have a supply curve that can be specified solely as a function of price because the monopoly's output depends on marginal revenue (which depends on the slope of the demand curve) and marginal cost.

The properties of the demand curve determine the monopoly *overcharge*, the amount by which the monopoly price, p_m , exceeds the marginal cost or competitive price, p_c , in Figure 4.2a. A relationship exists between the monopoly overcharge and the price elasticity of demand.

The elasticity of demand is a characteristic of the demand curve and is defined as the percentage change in quantity that results from a 1 percent change in price. If the elasticity of demand is very high (a large negative number), then the curve is said to be *elastic*. With a very elastic demand, a small price change induces a very large change in the quantity demanded. If the elasticity is low (a number between -1 and 0), the demand curve is *inelastic*, and a price change of 1 percent has relatively little effect on the quantity demanded.

Marginal revenue can be written as⁵

$$MR = p \left(1 + \frac{1}{\epsilon} \right), \quad (4.2)$$

where ϵ is the elasticity of demand. Thus, marginal revenue is positive if the demand curve is elastic ($\epsilon < -1$). It is negative if the demand curve is inelastic ($-1 < \epsilon < 0$). The elasticity of demand, in general, depends on not only the particular demand curve but also the point (the price and quantity pair) on the demand curve. For example, the elasticity of demand could decrease as price becomes lower.

By substituting Equation 4.2 for MR in Equation 4.1, we can write the profit-maximizing condition for the monopoly as:

$$\frac{p - MC}{p} = -\frac{1}{\epsilon}. \quad (4.3)$$

Thus, it sets $MR = MC$. Another condition for profit maximization is that the marginal revenue curve cut the marginal cost curve from above, as in Figure 4.2a. That is, the second-order condition must hold:

$$\frac{d^2\pi}{dQ^2} = \frac{dMR}{dQ} - \frac{dMC}{dQ} < 0.$$

A monopoly uses the same shut-down condition as does a competitive firm. In the short run, if price is below average variable cost, the monopoly stops producing.

⁵Differentiating revenue, $R = p(Q)Q$, with respect to Q , we find that the marginal revenue is

$$MR = p + \frac{dp}{dQ}Q = p \left(1 + \frac{dp}{dQ} \frac{Q}{p} \right) = p \left(1 + \frac{1}{\epsilon} \right),$$

where ϵ is defined as $(dQ/dp) (p/Q)$.

The left-hand side of Equation 4.3 is the **price-cost margin**: the difference between price and marginal cost as a fraction of price, $[p - MC]/p$. As the equation shows, the price-cost margin depends on only the elasticity of demand the monopoly faces. The price-cost margin is also called the **Lerner Index** of market power (Lerner 1934).

Equation 4.3 shows that the monopoly's price is close to MC when the demand is very elastic, and the price increasingly exceeds MC as the demand becomes less elastic. For example, if the elasticity of demand is -2 , price is twice marginal cost. If the elasticity is -100 (very elastic), price equals $1.01 MC$. The higher the elasticity of demand, the closer is the monopoly price to the competitive price. Therefore, the key element in an investigation of market power is the price elasticity of demand. Where the elasticity of demand is relatively inelastic, a monopoly markup may be substantial, as Example 4.1 illustrates.

Market and Monopoly Power

In contrast to a price-taking competitive firm, a monopoly knows that it can set its own price and that the price chosen affects the quantity it sells. A monopoly can set its price above its marginal cost but does not necessarily make a supracompetitive profit. For example, if a monopoly incurs a fixed cost, its profit may be zero (the competitive level) even if its price exceeds its marginal cost.

It is common practice to say that whenever a firm can profitably set its price above its marginal cost without making a loss, it has *monopoly power* or *market power*. One might usefully distinguish between the terms by using *monopoly power* to describe a firm that makes a profit if it sets its price optimally above its marginal cost, and *market power* to describe a firm that earns only the competitive profit when it sets its price optimally above its marginal cost. However, people do not always make this distinction, and generally use the two terms interchangeably, sometimes creating confusion.

EXAMPLE 4.1

Monopoly Newspaper Ad Prices

When the *Houston Post* shut down in April 1995, the managing editor of the sole surviving paper, the *Houston Chronicle*, received dozens of calls from concerned *Post* readers worried about one thing: Would the *Chronicle* pick up the *Post*'s comics? Local advertisers also were very concerned: What would happen to newspaper advertising prices?

Ad rates skyrocketed by nearly 62 percent from January 1995 (before the *Post* folded) to December 1996. The rate for a one-column inch ad in a daily paper rose from \$252.64 to \$409.00 per day, and Sunday rates jumped from \$294.84 to \$477.28. These rates increased by much more than readership, which rose 3.2 percent on weekdays and 23 percent on Sunday. Thus, a loss of competition resulted in a substantial increase in price.

Source: Iver Peterson, "New Realities of Life in a One-Paper Town," *New York Times*, December 30, 1996:C5.

The Incentive for Efficient Operation

Organized crime in America takes in over forty billion dollars a year and spends very little on office supplies.
—Woody Allen

The consequences of inefficient behavior are different for monopolies and competitive firms. An inefficient competitive firm may not be able to remain in business because it is unprofitable, but an inefficient monopoly can profitably remain in business. This observation has led some to conclude that the monopoly strives less hard to be efficient (called *x-inefficiency* by Leibenstein 1966) than does a competitive firm.

This argument is rejected by many economists who believe that monopolies, like other firms, prefer more to less. Monopolies want to maximize profits, and the only way a firm can do so is to minimize its costs at its chosen output level. Therefore, to postulate that monopolies want to maximize profits is to assume implicitly that they also minimize their costs. No firm—monopolistic or competitive—wants to throw money away. If improving the efficiency of operations increases profits, the firm should do it, whether it is a monopoly or a competitor.

A monopoly, however, may not have the same *ability* to produce as efficiently as a competitive firm. A firm in a market with many other firms can observe what other firms are doing. It can observe, for example, whether its own costs of production are above or below the market price. Because the market price reflects the efficiency of the other firms in the market, a competitive firm knows that it can improve its production efficiency if its costs of production are high relative to the market price. In contrast, a monopoly has no other firms to look at and may have no other standard by which to judge how efficiently it is operating. Therefore, a competitive firm may operate more efficiently than does a monopoly because it is more difficult for a monopoly to monitor internal efficiency than it is for a competitive firm.

Monopoly Behavior over Time

If demand is inelastic ($-1 < \epsilon < 0$), it is not possible to satisfy the profit-maximization condition of Equation 4.3. Thus, a monopoly never operates on the inelastic portion of its demand curve. If a monopoly were operating in the inelastic portion of its demand curve, it could increase its profits by raising its prices until it was operating in the elastic portion of its demand curve. In the inelastic portion of the demand curve, a 1 percent increase in the monopoly's price causes the quantity sold to fall by less than 1 percent, so that revenues increase. With reduced output, however, the monopoly's costs must fall, so that total profits must rise. Thus, if the monopoly is operating in the inelastic portion of the demand curve, it should keep increasing its price, obtaining ever more profits, until it is in the elastic portion of the demand curve.⁶

⁶What if there were no elastic portion of the demand curve? The monopoly would produce just a small amount of output, charge an infinite price, and make infinite profits. That this story is implausible underscores the empirical irrelevance of a monopoly's demand curve that is everywhere inelastic.

This observation, however, applies only in the context of a simple, timeless model. In actual markets, demand curves shift over time. As a result, a rational monopoly changes its price over time.

Consumers may have a more inelastic demand curve in the short run than in the long run. In the short run there are limitations on how fast consumers can substitute away from a product in the face of a price increase. Therefore, if a monopoly takes advantage of an inelastic portion of its short-run demand curve and raises its price, its consumers are more likely to substitute away from its product in subsequent periods. Thus, a monopoly may operate in the inelastic portion of its short-run demand curve to avoid long-run substitution.

The oil market provides an excellent example of the time it takes to substitute away from a product. When the Organization of Petroleum Exporting Countries (OPEC) raised the price of oil in the early 1970s, total consumption of energy changed very little in the first year. However, the quantity of oil demanded fell sharply over the next several years as consumers adjusted to the increased price and began to take energy-saving measures.

● The Costs and Benefits of Monopoly

A monopoly is socially reprehensible in the hands of others.

If a monopoly restricts its output and raises its price above marginal cost, society suffers a deadweight loss. We first examine why such behavior leads to a deadweight loss. Then we use our understanding of how monopolies arise to show that, in certain circumstances, there are benefits associated with monopolies. Indeed, in certain situations, monopoly may be preferable to competition.

The Deadweight Loss of Monopoly

In order to maximize its profit, a monopoly sets its output where its marginal revenue curve intersects its marginal cost curve, as Figure 4.2a shows. The gap between the monopoly's price and marginal cost represents the difference between the value (price) that buyers place on the product and the marginal cost of producing it. This gap is similar to the one caused by a tax on a competitive market (Chapter 3). In both cases, price and output differ from their competitive levels, and there is a deviation between the demand price (as given by the demand curve) and the supply price (as given by the marginal cost curve).

If consumers must pay a monopoly price p_m that is above the competitive price p_c , they lose consumer surplus equal to the sum of the monopoly profits and the deadweight loss in Figure 4.2a. The monopoly profit is less than the consumer surplus loss. Thus, society suffers a deadweight loss (the *DWL* triangle in Figure 4.2a) that equals the consumers' loss less the monopoly's gain. This *DWL* triangle is the area below the demand curve, above the marginal cost curve, and to the right of the equilibrium monopoly quantity.

Thus, both monopoly and an inefficient tax cause a deadweight loss. However, who keeps the transfer from consumers differs: Tax revenues go to the government, whereas

the monopoly keeps the monopoly profit. Even fairly small deadweight losses may be associated with a large redistribution of wealth, as the “monopoly profit” box in Figure 4.2a illustrates.

Many researchers have estimated the deadweight loss that monopoly imposes on the U.S. economy. In a pioneering paper, Harberger (1954) calculated that the deadweight loss is small: less than 0.1 percent of the gross national product (GNP: a measure of the value of all goods and services in our economy).⁷ Later researchers repeated these calculations based on different assumptions. Worcester (1973), for example, also finds that the *DWL* is small: 0.4 to 0.7 percent. Kamerschen (1966) estimates the *DWL* at 6 percent and Cowling and Mueller (1978) estimate that it is between 4 and 13 percent.⁸ Jenny and Weber (1983) find that the *DWL* in France is as high as 7.4 percent.

Rent-Seeking Behavior

The gods help those that help themselves.

—Aesop

Some researchers contend that the efficiency loss to society is much larger than the *DWL* triangle. They argue that an amount equal to some or all of the monopoly profits is also an efficiency loss.

Monopoly profits can be regarded as a transfer from consumers to the monopoly, just as tax revenues are a transfer of income from consumers to the government. By itself, a transfer of income does not affect efficiency. Only if the monopoly restricts output below competitive levels is there an efficiency effect.

However, Posner (1975) argues that the monopoly profits may also represent a loss to society to the extent that it creates incentives for a firm to use real resources to become a monopoly. For example, suppose that a firm can become a monopoly by persuading the government to pass a law that restricts entry into the market. The use of a firm's resources to hire lobbyists, lawyers, and economists to argue its case before legislators is a cost to society, because these resources could have been productively employed elsewhere.

If there is a positive monopoly profit, as in Figure 4.2a, a firm would be willing to spend an amount up to these profits in order to become a monopoly. Of course, the firm would like to spend as little as possible, but the opportunity to earn monopoly profit could create the incentive to use valuable resources up to the amount of monopoly profits in order to secure the monopoly.⁹ Because firms compete to earn the “rent”

⁷Stigler (1956) and Cowling and Mueller (1978) criticize Harberger's methodology on technical grounds.

⁸See, however, Masson and Shaanan (1984) for a critique of this last result.

⁹Whether the firm would dissipate the entire monopoly profit depends on the institutional details as to how the monopoly can be acquired (Fisher 1985).

(monopoly profits) from the monopoly, the expenditure of resources to attain government-created monopoly profits is called **rent seeking**.

If rent seeking occurs, the calculation of the deadweight loss from monopoly must include that part of the transfer that is dissipated by the firms seeking to become the monopoly. Thus, the cost of monopoly is greater than the *DWL* triangle that Harberger calculated: The loss equals the *DWL* triangle plus at least part of the monopoly profits.

Posner recalculates the deadweight loss from regulated and unregulated monopoly on the extreme assumption that the entire amount of monopoly profit is dissipated in rent-seeking activities. His estimates of deadweight loss as a percent of revenues exceed previous estimates. For example, Posner found deadweight losses of up to 30 percent of revenues for some of the industries he examined (such as motor carriers, physician services, and oil). His insight was that a great part of the loss to the economy from monopoly (or, more generally, noncompetitive pricing) is directly traceable to the existence of government institutions that insulate some firms from competition. If he's correct, the recent rescinding of many government regulations (see Chapter 20) will provide sizable benefits to society.

Monopoly Profits and Deadweight Loss Vary with the Elasticity of Demand

Monopoly profits and the *DWL* triangle depend on the shape of the demand curve. We illustrate how monopoly profits and deadweight loss vary with the elasticity of demand with a linear demand curve,

$$p = a - bQ.$$

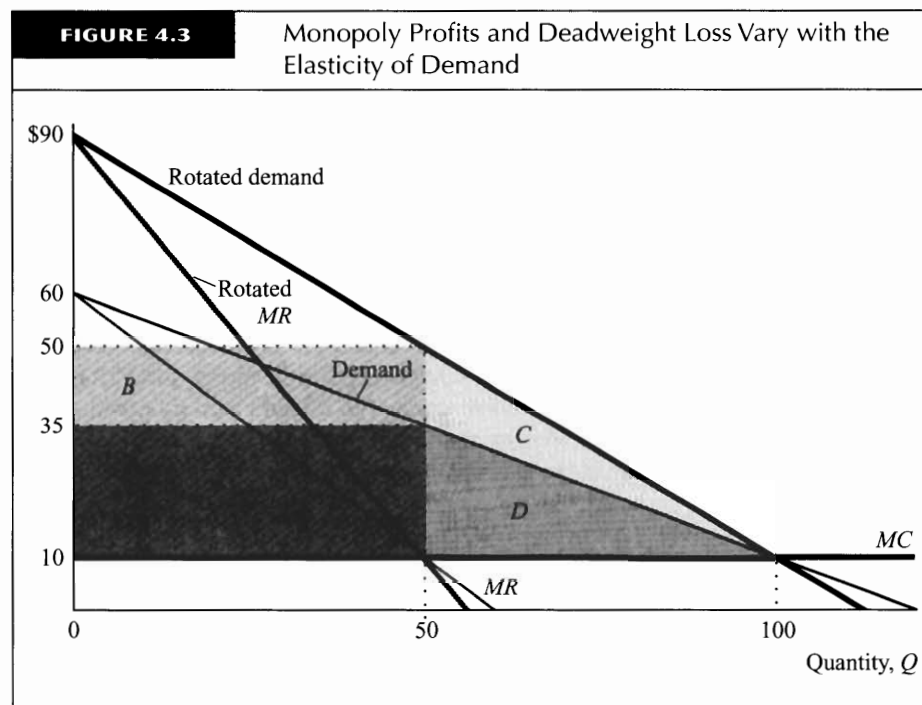
The light demand curve in Figure 4.3 is for $a = \$60$ and $b = 0.5$. Given a constant marginal and average cost, $MC = AC = \$10$, the monopoly sells $Q_m = 50$ units at $p_m = \$35$, where the elasticity of demand is -1.4 . The monopoly's profit is Area $A = \$1,250$, and the deadweight loss is area $D = \$625$.¹⁰

We now rotate the demand curve so as to vary the elasticity of demand. The demand curve is rotated around the point where it crosses the *MC* line, at 100 units. That is, for all the demand curves examined, if price were set efficiently at $MC = \$10$, consumers would buy 100 units. Because the demand curve is linear, the marginal revenue curve is also linear and crosses the horizontal *MC* line at half the distance that the demand curve does. Thus, the profit-maximizing monopoly equilibrium quantity of 50 units is unchanged as we rotate the demand curve.

¹⁰Let $t (= [p_m - p_c]/p_c)$ be the monopoly markup above the competitive price. For small t , the monopoly *DWL* triangle can be approximated as

$$-1/2 t^2 R \epsilon,$$

where R would be the revenues if the product were sold at the competitive price ($p_c Q_c$) and ϵ is the elasticity of demand. *DWL* does not necessarily rise as the absolute value of ϵ increases because t is inversely related to ϵ , and as t changes, so does R . Holding R constant, *DWL* falls as the absolute value of ϵ increases.



The thick blue demand curve in Figure 4.3 shows a demand curve that has been rotated so that its intercept with the price axis, a , is \$90, which is higher than the original \$60 intercept. The thick blue demand curve intercepts the price axis at \$90. The monopoly sells the same quantity as before, $Q_m = 50$, but at a higher price, $p_m = \$50$, so that the demand elasticity falls (in absolute value) from -1.40 to -1.25 (becomes less elastic), as Table 4.1 illustrates. The monopoly profit rises to $A + B = \$2,000$, and the deadweight loss increases to area $C + D = \$1,000$.

TABLE 4.1 Monopoly Profits and Deadweight Loss Vary with the Price Elasticity of Demand

Intercept of Demand Curve with Price Axis	Elasticity of Demand at $Q = 50$	Monopoly Price	Deadweight Loss	Monopoly Profit
\$ 30	-2.00	\$20	\$ 250	\$ 500
60	-1.40	35	625	1,250
90	-1.25	50	1,000	2,000
120	-1.18	65	1,375	2,750
150	-1.14	80	1,750	3,500

As the demand curve becomes less elastic at the monopoly equilibrium, people are less willing to do without this good: An increase in price causes the quantity they purchase to fall by less than if demand were more elastic. The monopoly, realizing this opportunity exists, increases its equilibrium price and earns a larger monopoly profit. As the demand curve becomes steeper at a given quantity (demand is more inelastic), the deadweight loss increases.

The Benefits of Monopoly

The welfare harms from monopoly may be offset by several benefits. These benefits are ignored in the static analysis above where we calculated deadweight losses. For example, the prospect of receiving monopoly profits may motivate firms to develop new products, improve products, or find lower-cost methods of manufacturing. Were it not for the quest to obtain monopoly profits, firms might innovate less.

The benefit of monopoly is most clearly recognized in research and development (see Chapter 16). If a firm succeeds in developing a new product, it can obtain a patent that prohibits other firms from using the patented technology for a fixed number of years—currently 20 years in the United States. Were it not for the patent, the innovative firm might discover that, within a matter of weeks, other firms had copied the new product. The innovative firm would then receive no more than the competitive level of profits and would not recover its expenditures on research and development. The firms that copied the product would have no research and development expenditure to recover. The ability of other firms to copy a new product removes the innovating firm's incentive to invest in research and development. The patent system attempts to deal with this problem by granting the innovating firm the sole property right to commercially exploit its innovation.

Naturally, if monopoly had no offsetting benefits, competition would be preferable. For example, if all firms in a competitive market decide to merge, and if the merger does not lead to a more efficient market, then the only result is the creation of a monopoly. As long as new entry takes time, the firms could price above their marginal cost. Because there is no benefit from this action, such behavior should be discouraged. One responsibility of the Department of Justice and the Federal Trade Commission is to scrutinize each merger carefully to make sure that its effect is not simply to raise prices to consumers.

Creating and Maintaining a Monopoly

There are several ways in which a firm may become and remain a monopoly. One possibility is that all the firms *merge* (combine into a single firm) or act in concert as a monopoly would. We address these possibilities in detail in Chapter 5 and in Example 4.2. Another possibility is that the firm takes strategic actions that prevent entry by other firms, as we discuss in Chapter 11 and in Example 4.3. Here, we examine three other reasons why a firm is able to create and maintain a monopoly:

EXAMPLE 4.2***Monopolizing by Merging******United States***

In 2001, the Federal Trade Commission (FTC) accused the Hearst Corporation of illegally acquiring a monopoly over medical drug databases that are used by pharmacies and hospitals. Hearst bought Medi-Span, which was the only major competitor for Hearst's database company, First DataBank, according to the FTC. The FTC went on to contend that Hearst withheld information necessary for its premerger antitrust review. After Hearst acquired Medi-Span, it raised prices, doubling some and tripling others, according to the FTC and Express Scripts, a pharmacy-benefit management company. In its settlement with the FTC, Hearst agreed to return \$19 million to customers. Later, Hearst paid more than \$26 million to Express Scripts and other class-action plaintiffs in a private antitrust suit in 2002.

South Africa

South African Breweries controls 98 percent of South Africa's beer sales, with its 14 brands, including Castle, Lion, Heineken, Guinness, Amstel, and Carling Black Label. It was formed by a merger of two major competitors in 1979 because South Africa had virtually no antitrust laws. A company spokesman claims that the firm has little market power because the market is "fully contestable" with no legal barriers to entry. The firm's control of distribution channels may be responsible for its ability to maintain its high market share.

Sources: "FTC Accuses Hearst of Creating Monopoly," *San Francisco Chronicle*, April 15, 2001:D2; "Hearst Settles Dispute with FTC," *Milwaukee Journal Sentinel*, December 15, 2001:D1; Peter Shinkle, "Express Scripts Drops Antitrust Suit vs. Hearst; Maryland Heights Company with Share in FTC Settlement," *St. Louis Post-Dispatch*, May 23, 2002:C11; Donald G. McNeil, Jr., "In South African Beer, Forget Market 'Share'," *New York Times*, August 27, 1997:C1, C4; Bernard Simon, "Private Sector; An Old School Brewer for Miller," *New York Times*, February 2, 2003:3.2.

The firm may have special knowledge, the government may protect it from entry, or the market may only be large enough for a single firm to produce profitably.

Knowledge Advantage

A firm may be a monopoly because only it knows how to produce a certain product or it can produce the product at lower cost than other firms. A firm may have special knowledge that enables it to produce a new or better product that others cannot imitate. The firm may try to keep secret its special knowledge so as to prevent rivals from imitating it (see Example 4.4). A firm with an important secret faces a downward-sloping demand curve for its product and does not fear the entry of rival firms or the introduction of products that are close substitutes.

EXAMPLE 4.3***Controlling a Key Ingredient***

In 2000, the attorneys general of 33 states and the Federal Trade Commission settled a lawsuit with Mylan Labs (and its suppliers) for \$100 million. The suit contended that Mylan Labs cornered the market on the active ingredients for two drugs used to treat Alzheimer patients and then raised the price of the drug Clorazepate more than 3,000 percent (from about 2¢ a tablet to over 75¢) and increased the price of the drug Lorazepam more than 2,000 percent (from about 1¢ a tablet to over 37¢).

Source: www.state.ia.us/government/ag/mylan.htm.

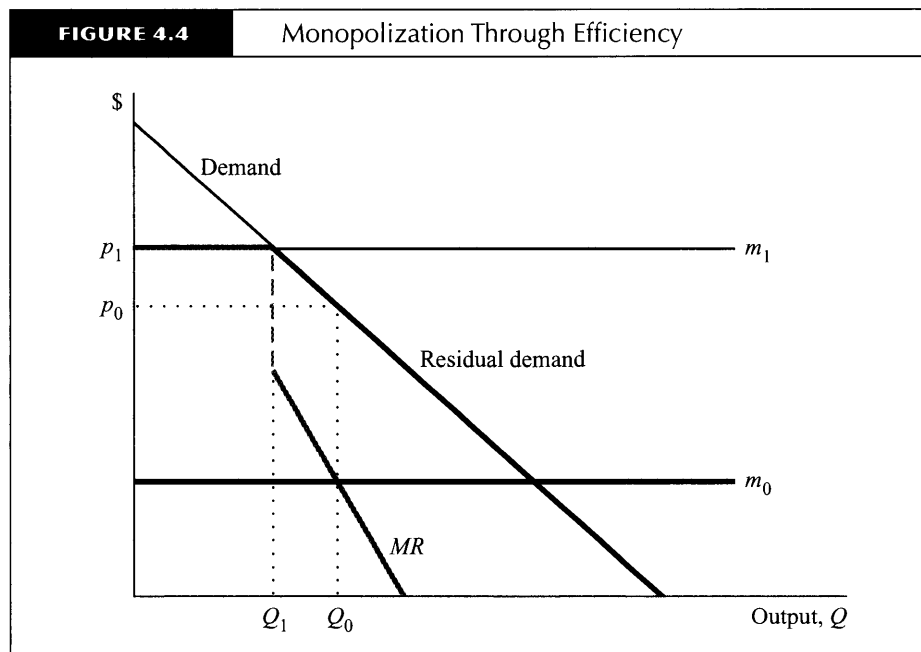
Similarly, a firm may have special knowledge about production techniques that enable it to produce the same product at lower cost than other firms, which may be unable to discover the production technique of the efficient firm. We illustrate this possibility in Figure 4.4. Initially all the firms in a competitive market have a constant marginal cost, m_1 , so the equilibrium price, p_1 , equals m_1 and the equilibrium quantity is Q_1 . One firm discovers a new production technique that it can keep secret and that lowers its marginal costs from m_1 to m_0 . It faces a *residual demand curve* (the unmet demand after all other firms sell as much as they want at a given price) that is horizontal at p_1 (equal to m_1) up to Q_1 because many firms can produce and sell at price m_1 . Beyond Q_1 (prices less than p_1), the residual demand curve coincides with the market demand curve because below p_1 no other firm can profitably produce.

If m_0 is close to m_1 , the firm may maximize its profit by selling at a price equal to p_1 . However, in Figure 4.4, m_0 is enough less than m_1 that the profit-maximizing monopoly price is less than m_1 but above m_0 . Because the residual demand curve has a kink in it at Q_1 , the corresponding marginal revenue curve is discontinuous at the output, Q_1 . The marginal revenue curve is horizontal where the residual demand curve is horizontal and slopes down where the residual demand curve is downward

EXAMPLE 4.4***Preventing Imitation—Cat Got Your Tongue?***

Why are violin strings called *catgut* when they are really made of sheep intestines? An old Roman named Erasmo (c. 130 AD) started making strings for musical instruments out of sheep intestine. The demand grew. Because it was considered extremely bad luck to kill a cat, Erasmo identified his product as *catgut* so nobody would imitate it and ruin his monopoly.

Source: L. Boyd, "Grab Bag," *San Francisco Chronicle*, October 27, 1984:35.



sloping. To maximize its profit, the firm with the secret process produces Q_0 units of output where its marginal revenue curve equals its marginal cost curve. The firm sets its price at p_0 , which is less than $p_1 = m_1$, so no other firm remains in the market.

Government-Created Monopolies

A firm may be a monopoly because the government protects it from entry by other firms. For example, suppose a firm invents a new product and realizes that imitation *is* possible technically. In most countries, the original innovating firm can obtain legal protection to prevent entry for some period of time. The law on intellectual property, in particular the patent law, grants a legal monopoly to a firm that has discovered a new product or technique. A firm can obtain a patent (see Chapter 16) on a new product that prevents any other firm from copying its product and competing with it for a fixed period of years.

Aside from the patent laws, other types of government (or government sanctioned) restrictions on entry can serve to create and maintain monopolies. Generally, government restrictions on entry allow at least a few firms to produce, but they prevent the normal competitive forces from driving price and profits down to competitive levels (see Example 4.5).

For example, in many cities, one must purchase medallions, of which only a fixed number are sold by local authorities, to operate a taxicab. The United States, by granting exclusive (monopoly) rights to portions of the electromagnetic spectrum, gave broadcast television stations at least \$40 billion in present-value terms for the first 30 years of television (Isé and Perloff 1997).

EXAMPLE 4.5*Protecting a Monopoly*

An 1872 law established the U.S. Postal Service (USPS) monopoly on mail delivery. In 1971, the USPS started an express mail service. A 1979 amendment to the 1872 law broke the agency's monopoly on urgent mail, establishing as the definition of *urgent*, mail that must arrive by noon the next day or lose its value. However, the USPS has the right to decide what is urgent and what is not.

How serious was the Postal Service's competition on express mail? By 1994, the USPS's share of the express mail market had fallen to just under 15 percent. Worse, to the horror of postal officials, the federal government contracted with Federal Express for next-day delivery of government parcels at a price of \$3.75, well below the Postal Service's overnight Express Mail rate of \$9.95.

The USPS fought back. From 1990 through 1993, the Postal Service fined 21 companies for violating the USPS's legal monopoly on mail delivery, collecting more than \$542,000 in fines from these companies that sent "nonurgent" mail by private couriers such as Federal Express, UPS, and DHL.

For example, an Atlanta-based credit-reporting company, Equifax Inc., was assessed a \$30,000 penalty, making up the loss to the Postal Service for routine business mail it sent by express services. Postal officials say they recovered \$4 in lost revenue for every \$1 spent on enforcement.

In 1994, the USPS issued a postal inspectors' audit that found that five federal agencies—the General Services Administration (GSA) and the departments of Agriculture, Health and Human Services, Treasury, and Energy—routinely infringed on the USPS's monopoly on first-class mail by using Federal Express to ship materials that were not time sensitive.

The report warned that the agencies, which accounted for one-third of the 4.3 million government packages moved by Federal Express in the first two years of the contract, "are incurring a substantial liability for postage"—the revenue that would have otherwise gone to the Postal Service. The USPS did not demand any payments for the postage, but postal officials pressured the GSA to train federal mailroom personnel as to what kind of materials they can legally send by Federal Express.

Armed with news reports of USPS fines on private firms and pressure on federal agencies, outraged private companies went to Congress for legislation ending the Postal Service's practices. Smarting from bad publicity and congressional pressure, the USPS announced that it would cease its practice of raiding businesses to check up on their use of commercial overnight delivery services, and stopped complaining about federal agencies. Nice try, though.

Sources: Michael A. Goldstein, "Can the U.S. Postal Service Market Itself to Success?" *Los Angeles Times Magazine*, December 22, 1996:14; Bloomberg News, "UPS Aims to Curb Postal Service Monopoly," *The Dallas Morning News*, April 14, 1998:9D; Bill McAllister, "Must It Get There Overnight?: Agencies Improperly Bypassing Postal Service, Inspectors Report," *Washington Post*, January 12, 1994:A17; "Private Couriers and Postal Service Slug It Out," *New York Times*, February 14, 1994:D2.

Until recently, U.S. states required someone wishing to build an in-patient medical facility to obtain a certificate of need by demonstrating that a new facility was needed. Using these laws, an early entrant could make entry by potential competitors difficult. In part because of these laws, Community Psychiatric Centers, a chain of psychiatric hospitals in the United States and Britain, had annual earnings growth of 15 to 30 percent between 1969 when it went public and 1985.¹¹

Similarly, trade barriers can be used to prevent entry. For example, in 1992, the Ontario government agency that monopolizes the sale of beer in that province, the Liquor Control Board of Ontario, announced a ban on American beer imports. Similarly, China places a 230 percent tariff (tax on foreign products) on foreign cigarettes to protect the China National Tobacco Corporation, which sells 1.75 trillion of the 5 trillion cigarettes sold throughout the world and accounts for 12 percent of the revenue of the Chinese government.¹²

Natural Monopoly

In some markets, it is efficient for only one firm to produce all of the output. When total production costs would rise if two or more firms produced instead of one, the single firm in a market is called a **natural monopoly**.

A firm is a natural monopoly if it can produce the market quantity, Q , at lower cost than can two or more firms. Let q_1, \dots, q_k be the output of the k (≥ 2) firms in a market that produce an identical product so that total market output equals the sum of the firms' output: $Q = q_1 + \dots + q_k$. If each firm has a cost function $C(q_i)$ and one firm can produce Q at lower cost than the sum of the k firms,

$$C(Q) < C(q_1) + C(q_2) + \dots + C(q_k),$$

then the least expensive (most efficient) way to produce is to have one firm produce all Q units. A cost function is said to be *subadditive* at Q if this inequality holds, so subadditivity is a necessary condition for the existence of a natural monopoly (Sharkey 1982; Baumol, Panzar, and Willig 1982).

A natural monopoly often has falling average costs and constant or falling marginal costs in the region in which it operates. A strictly decreasing average cost curve implies subadditivity (though the opposite does not necessarily follow).

Suppose that the average cost curve of a natural monopoly is downward sloping, and that the firm can produce 100 units at an average cost of \$10 per unit. The firm's total cost of producing that many units is \$1,000. Now suppose that a second firm with identical costs enters the market. If each of these two firms produces 50 units, their average cost of production is higher than before because the average cost curve is downward sloping. If their average cost is \$15 per unit, for example, their combined

¹¹See www.aw-bc.com/carlton_perloff "Model of Insanity."

¹²Glenn Collins, "U.S. Tobacco Industry Looks Longingly at Chinese Market, but in Vain," *New York Times*, November 20, 1998:A10.

total cost of producing 100 units is \$1,500. Thus, a single firm can produce 100 units at lower cost than can two firms.

It is often argued (but may not be true) that electrical, gas, telephone, and cable television are natural monopolies. There is a relatively high fixed cost for running an electric power line or a phone line to a home or firm, but constant or falling marginal costs of supplying the service. As a result, marginal cost is constant or falls, and average cost falls as output increases.¹³

If production is characterized by economies of scale everywhere, then average cost declines as output increases, and it is always less costly for one firm to produce any given output than for several firms to produce that output. Therefore, when average cost falls with output, there is a natural monopoly. A natural monopoly can occur even if average cost is not declining everywhere with output. For example, if a U-shaped average cost curve reaches a minimum at an output of 100, it may be most efficient for only one firm to produce an output of 101 even though average cost is rising at that output. Therefore, economies of scale are a sufficient but not a necessary condition for natural monopoly.¹⁴

● Profits and Monopoly

Many people associate high profits with monopoly or too little competition, normal profits with competition, and losses with excessive competition. Although each of these beliefs has some element of truth, none is correct. We now show why these beliefs do not hold in general by answering three questions: (1) Is anyone who earns positive profits a monopoly? (2) Does a monopoly always earn positive profits? (3) Should the government allow mergers that create monopoly in a market that was suffering short-run losses?

Is Any Firm That Earns a Positive Profit a Monopoly?

Although a monopoly may earn positive profits, it does not follow that any firm that earns a positive profit is a monopoly. The previous chapter discusses the possibility that certain scarce resources, such as land, can earn rents. For example, a wheat farmer who owns particularly productive land earns a large profit. This profit is attributable to the land that is owned and should properly be called a rent. The farmer behaves competitively, taking price as given and operating where price equals marginal cost. This

¹³The empirical literature, however, leaves some doubt as to whether many utilities exhibit increasing returns to scale, which implies downward-sloping marginal and average cost curves. Moreover, showing that there are scale economies in one range of output is not sufficient to demonstrate that a firm is a natural monopoly (that is, the cost function is subadditive). See, for example, Fuss and Waverman (1981) and Evans and Heckman (1982a, 1982b). Shin and Ying (1992) argue that local telephone exchange carriers were not natural monopolies prior to deregulation. Friedlaender (1992) finds evidence of substantial returns to scale for railroads.

¹⁴In Chapter 20, we examine how governments regulate natural monopolies and the conditions under which other firms will try to enter a market with a natural monopoly.

farm is a competitive firm; rents on factors of production do not indicate a monopoly. As long as output is not restricted so that price equals marginal cost, there is no market power. Scarce resources can command very high prices and those who own those resources benefit. For example, star athletes earn high salaries (rents) even though they are not monopolies that restrict output.

Does a Monopoly Always Earn a Positive Profit?

Although a monopoly earns a larger profit than a competitive firm would, it is not true that a monopoly always earns a positive profit. In the short run, a monopoly can make losses, just as a competitive firm can. A monopoly that faces a sudden decline in demand may continue to operate even though it makes a negative short-run profit (its price is less than its average cost) if its price is above its average variable cost. Losses in a market do not imply that it is competitive. In the long run, when there are no sunk costs, no firm continues to operate if there are only losses in the market.

As in competition, the length of time that losses will be earned by a monopoly depends on how long the short run lasts—how long it takes for the plant and equipment to wear out, forcing a decision on whether to replace them. In some markets, the short run may be very long. For example, railroad tracks can last for years or possibly decades. Therefore, one might expect that a monopoly railroad could earn a negative profit on its investments for a long time before deciding to exit the market.

Briefly, in the long run, a competitive firm makes zero economic profit, whereas a monopoly makes a zero or positive profit. In the short run, both competitive firms and monopolies may make losses or profits.

Are Monopoly Mergers to Eliminate Short-Run Losses Desirable?

A merger of firms into a monopoly can eliminate competition and allow the merged firm to exercise market power and raise price so that the losses are eliminated. Firms in a market where all firms are losing money often argue for a merger for this reason (see Example 4.6). This motivation for merger appears to have a certain logical appeal—if

EXAMPLE 4.6

EU Allows Merger to Eliminate Losses

In 2003, the European Union allowed Rupert Murdoch's News Corporation to merge Telepiu, which had two-thirds of all pay-TV subscribers in Italy, with its own Italian pay-TV firm, Stream, to create a new firm called Sky. EU Competition Commissioner Mario Monti conceded that his decision "will create a quasi-monopoly on the Italian market." He justified his actions by saying that a weak business environment allowed room for only one firm to survive in this market, as both Stream and Telepiu had been losing money.

Source: Raf Casert, "EU Commission Allows Murdoch's News Corp. to Forge 'Quasi-Monopoly' in Italian Pay TV," Associated Press, April 2, 2003.

the merger eliminates the losses, perhaps it is efficient for the merger to occur. However, such a merger harms society!

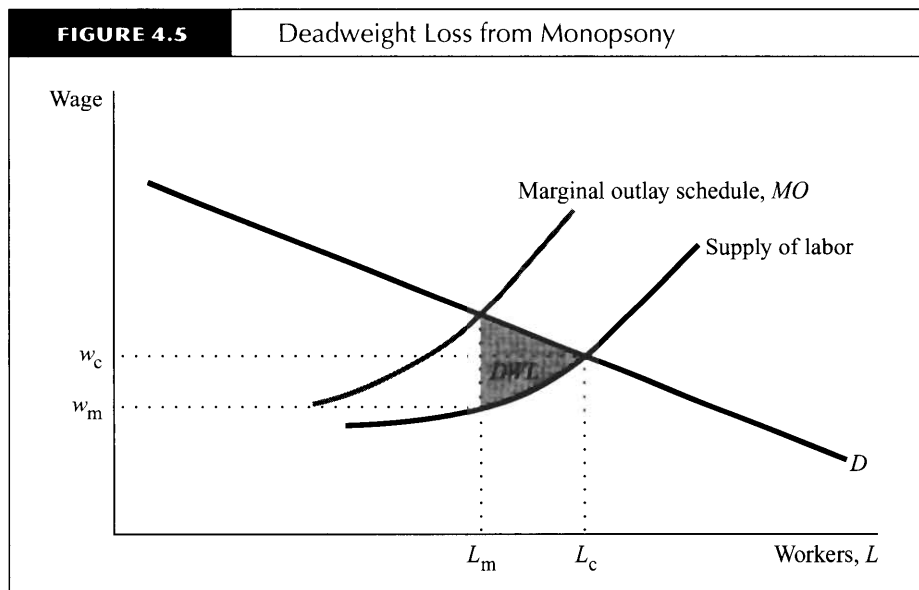
If a merger enables firms to set a price in the short run that is greater than the level at which they would have priced had they remained competitive, then the merger imposes a deadweight loss on society. The existence of sunk costs in the short run that cause short-run losses cannot be eliminated by merging firms. The merger only changes the amount of competition that firms face. Because the merger does not eliminate sunk costs, it is inefficient to allow firms to form a monopoly and thus allow the price to rise.

Monopsony

A single buyer in a market is called a **monopsony**. A monopsony's decision on how much to buy affects the price it must pay (just as a monopoly's choice of output affects the price it receives). The monopsony decides how much to purchase by choosing a price-quantity pair on the market supply curve. Monopsony is the flip side of monopoly. Both a monopoly and a monopsony recognize that their actions affect the market price.

A monopsony determines how much to buy in much the same way that a monopoly determines how much to produce. A monopsony buys more of the good as long as the value of the extra consumption as given by its demand curve equals or exceeds its marginal cost of consuming one more unit.

If there is a competitive labor market, each firm takes the wage rate as given, and the marginal cost of hiring one more worker is simply the wage rate. Now suppose there is only one local employer (buyer of labor services): a monopsony. In Figure 4.5,



it faces an upward-sloping supply curve for labor. In order to hire an extra worker, the monopsony must not only pay that worker a slightly higher wage rate but also pay *all* its other workers a slightly higher wage rate, because only by raising the wage can extra labor be induced into the marketplace.

If the monopsony must raise its wage from, say, \$5 to \$6 to induce that last individual to work for the firm, the monopsony's extra cost of hiring the additional worker is not just \$6; it is \$6 plus the \$1 increase in wages that must be passed along to each of its original workers. If it originally had 100 workers, its total wage bill rises from \$500 to \$606: an increase of \$106. The monopsony recognizes that *its* marginal cost of hiring the additional worker is \$106 rather than \$6 and takes that into account in deciding whether to hire the additional worker. The monopsony hires an extra worker only if the marginal benefit as given by its labor demand curve exceeds its marginal cost of hiring an additional worker.

The marginal cost to a monopsony of buying additional units (hiring additional workers) is described by a **marginal outlay schedule**, which is analogous to a marginal revenue curve. As Figure 4.5 illustrates, the marginal outlay schedule lies above the upward-sloping supply curve because the monopsony must raise the wage for all its workers to hire an extra worker. A profit-maximizing monopsony hires L_m workers, where its marginal benefit, as given by its demand curve, equals its marginal outlay. Because the marginal outlay curve lies above the supply curve, the monopsony hires fewer workers, L_m , than would a competitive market, which hires L_c workers (determined by the intersection of the demand curve and the supply curve). In other words, a monopsony restricts output just as a monopoly does.

The monopsony wage rate, w_m , is below the competitive wage rate, w_c . Using a definition analogous to the one for market power, we can define monopsony power as the ability to profitably set wages (or other input prices) below competitive levels. At the monopsony solution (L_m , w_m) in Figure 4.5, there is a gap between the demand curve and the supply curve. A gap between the demand curve (which represents the marginal benefit to society of consumption) and the supply curve (which represents the marginal cost to society) reflects a loss in efficiency. The monopsony deadweight loss triangle (Figure 4.5) is analogous to the deadweight loss that results from monopoly (Figure 4.2a).

Most labor economists believe there are few monopsonized labor markets in the United States. Example 4.7 identifies one such market. The most frequent examples given of monopsony in the labor market concern single-company towns, local employment markets, and sports leagues. For example, a major league baseball player can work in the United States only if he plays for a team that belongs to either the American or National Leagues. Collectively, these teams are the sole buyer in the United States for the services of a major league baseball player. To the degree that the teams agree not to compete for players, they gain monopsony power. To offset such monopsony power, baseball players can form a union to obtain monopoly power in selling their labor services.

Monopsony is most likely in markets where resources are specialized to a few uses. Moreover, even if resources are initially specialized to one use, as with a piece of custom-designed machinery (or a plant in a specific location serving a single buyer),

EXAMPLE 4.7*Priest Monopsony*

Newspapers repeatedly report a “shortage” of Catholic priests, citing sources both inside and outside the church hierarchy. Between 1960 and 2000, the number of priests declined 13 percent even as 55 percent more Catholics had joined their local parishes. Strikingly, other religious denominations are not suffering from a “shortage of clergy.”

Why the difference between churches? Is it due to changing tastes among potential priests or some other factor? Daniel Condon (2002) attributes the difference to the fact that the Catholic Church exercises monopsony power, whereas other churches and synagogues permit an active, competitive labor market for clergy. Individual Catholic parishes do not compete for clergy, who are instead assigned by the central church authority (diocese). Wages vary little across parishes, although priests in wealthy parishes may receive larger fees for performing wedding and funeral services.*

Condon estimates that Catholic priests earn 41 percent less than non-Catholic officiants, controlling for education, experience, location, and whether they are provided rent-free housing. He concludes (p. 894) that the true differential is “more pronounced when one considers that Catholic clergy have a condition of employment (celibacy) that would require additional monetary compensation for most.”

*Academics at a Catholic university may face an even greater problem. For them, it's publish or perish.

monopsony may not persist in the long run. The reason is that no one will make new custom-designed machinery (or new investments in a plant) for a specific buyer if they earn a depressed return compared to what they can earn from making other machines (or building a plant elsewhere). In other words, few resources are specialized in the long run, and therefore it is unlikely that monopsony can persist in the long run.

Another way to explain the preceding point is as follows. If resources are not specialized to a particular market in the long run, then the long-run supply curve tends to be flat (highly elastic). As Chapter 3 explained, a flat long-run supply curve is most likely to occur when the market in question uses only a relatively small fraction of the total consumption of its inputs. Long-run monopsony power is impossible if the long-run supply curve is flat because price cannot be lowered below the competitive price.

If the long-run supply curve is flat, there may not be any monopsony power even in the short run. Suppose that before a firm enters a market, it has many alternative uses for the resources it owns. After it enters the market, it specializes its machines so that it has very few alternative uses for its assets. Suppose that it will only enter a particular market if it receives \$10 per unit of output (which is the long-run average cost). The sole buyer, the monopsony, agrees to pay \$10. After the firm enters, it is committed, at

least for some time, in the sense that its machines are specialized to this particular market. If the monopsony lowers its price to \$9, it may not pay for the firm to exit immediately. But the firm will not replace the specialized machines when they wear out, and the monopsony may eventually have no one willing to supply the product. Even if the buyer again promises \$10 per unit to induce a supplier to enter, no firm would believe the buyer in light of its previous behavior. So, for a buyer that is concerned about a long-run source of supply, it may not pay to exercise short-run monopsony power.

Dominant Firm with a Competitive Fringe

*Where does the gorilla sleep?
Anywhere the gorilla wants to sleep.*

What happens to a monopoly if other, higher-cost firms enter its market? Or, similarly, what happens if a lower-cost firm enters a market with many price-taking, higher-cost firms? After entry, the lower-cost firm has a relatively large share of the market. If one firm is a price setter and faces smaller, price-taking firms, it is called a **dominant firm**. It typically has a large market share. The smaller, price-taking firms, called **fringe firms**, each have a very small share of the market, though collectively they may have a substantial share of the market.

There are several industries in which one firm has a large share of the industry sales. For example, Kodak's share of the photographic film business has been estimated at 65 percent.¹⁵ Hewlett-Packard is estimated to have 59 percent of laser printer sales.

We begin by discussing what makes a firm dominant. We then analyze how entry limits a dominant firm's market power. We examine two extreme cases. In the first, entry by other firms is impossible. In the second, entry by competing fringe firms can occur instantaneously. The analysis shows that a dominant firm's price-setting behavior depends on the ease of entry by fringe firms.

We draw two main conclusions. First, it is generally not in a profit-maximizing dominant firm's best interest to set its price so low that it drives all competitive-fringe firms out of the market. Second, the presence of competitive-fringe firms or the threat of entry by additional firms may force a dominant firm to set a price lower than the price a monopoly would set (see Example 4.8).

If a sufficiently large number of price-taking firms can enter the market, a dominant firm cannot continue to charge a price higher than the minimum average cost of these new firms. Indeed, if potential entrants' costs are as low as the dominant firm's, the dominant firm eventually has no more market power than any other firm.

¹⁵A firm's share of sales in an industry depends crucially on how the industry is defined, and hence is often controversial, especially in court proceedings.

EXAMPLE 4.8*Price Umbrella*

It is often asserted that a dominant firm provides a *pricing umbrella* for smaller firms. As long as competing firms price at or below the level of the dominant firm, they will be able to find buyers. If their products are inferior (say because they are risky to use for legal reasons), the fringe firms have to set their prices substantially below the dominant firm's.

In many countries, phone monopolies charge rates that are more than twice those in the United States, where competition has kept rates relatively low. This price difference causes problems for the monopolies.

"Callback" services offer some customers a way to evade paying high monopoly prices. A callback service provides a "trigger" number connected to a computer in the United States. The customer calls that number using the monopoly service and hangs up before the phone is answered, paying nothing for the incomplete call. The computer calls the customers back and offers an American dial tone, which can be used to place a call anywhere in the world for rates well below the monopoly price. In some cases, the callback rates are less than the price of a local call. Hundreds of American companies provide these services, and the rate of use has grown exponentially over time. Ghana's monopoly is reported to lose \$1 million each week to callback and Internet services.

To protect local monopolies, governments in many countries—including Argentina, Canada's Northwest Territories, China, Malaysia, Saudi Arabia, South Korea, and Uganda—try to stop these services. The U.S. operators believe they are beyond the reach of local laws. For example, when Uganda blocked all calls to the Seattle, Washington, area code where one service, Kallback, is based, the company routed the calls through a different area. When other countries tried to identify and block the services by picking up the touch-tone beeps used to complete calls, Kallback added a voice-recognition system. As a firm spokesman said, "It's a cat and mouse game. It's kind of fun."

Source: "Don't Call US," *The Economist*, 338(7947), January 6, 1996:55; www.kallback.com; "Telecom Loses \$1m a Week, Communications Experts Say," *Ghanaian Chronicle*, February 7, 2003.

Why Some Firms Are Dominant

All animals are equal, but some animals are more equal than others.

—George Orwell

Why do some firms gain substantial market power, while others do not? At least three possible reasons are sufficient to create a dominant firm-competitive fringe market structure.

The first reason is that *dominant firms may have lower costs than fringe firms*. There are at least four major causes of lower costs:

- A firm may be more efficient than its rivals. For example, it may have better management or better technology that allows it to produce at lower costs. Such a technological advantage may be protected by a patent.
- An early entrant to a market may have lower costs from having learned by experience how to produce more efficiently.
- An early entrant may have had time to grow large optimally (in the presence of adjustment costs) so as to benefit from economies of scale. By spreading fixed costs over more units of output, it may have lower average costs of production than a new entrant could instantaneously achieve.
- The government may favor the original firm. The U.S. Postal Service does not pay taxes or highway user fees, which reduces its cost relative to that of competing package delivery services.

A second important reason is that *a dominant firm may have a superior product* in a market where each firm produces a differentiated product. This superiority may be due to a reputation achieved through advertising or through goodwill generated by its having been in the market longer.

A third reason is that *a group of firms may collectively act as a dominant firm*. As Chapter 5 shows, groups of firms in a market have an incentive to coordinate their activities to increase their profits. A group of firms that explicitly acts collectively to promote its best interests is called a *cartel*. If all the firms in a market coordinate their activities, then the cartel is effectively a monopoly; if only some of them do so, then the group acts as a dominant firm facing a competitive fringe of noncooperating firms.

One example of a dominant firm is the cartel consisting of Philippine coconut-oil-producing firms that act in concert but face a fringe of firms in other countries that act as price takers. With nearly four-fifths of the world's export market, the Philippine cartel has dominant-firm market power with a Lerner Index of 0.89 (Buschena and Perloff 1991).

Whether a dominant firm can exercise market power in the long run depends crucially on the number of firms that can enter the market, how their production costs compare to those of the dominant firm, and how fast they can enter. We now examine the dominant firm-competitive fringe model under two alternative extreme assumptions about the ease of entry.

The No-Entry Model

Consider a market with a dominant firm and a competitive fringe in which no additional fringe firms can enter. Two key results emerge from an analysis of this model: (1) It is more profitable to be the *gorilla* of a market than a mere fringe firm. (2) The existence of the fringe limits the dominant firm's market power—that is, it is more profitable to be the only firm in a market (a monopoly) than merely a dominant firm.

Assumptions. Five crucial assumptions underlie this no-entry model:

1. *There is one firm that is much larger than any other firm because of its lower production costs.* Although a market may be characterized by a small group of relatively large firms rather than a single dominant firm, we concentrate on the case of the single dominant firm for simplicity.
2. *All firms, except the dominant firm, are price takers,* determining their output levels by setting marginal cost equal to the market price (p).
3. *The number of firms (n) in the competitive fringe is fixed: No new entry can occur.* That is, the dominant firm knows that it can raise the market's price without causing new firms to enter the market or existing firms to build additional plants.
4. *The dominant firm knows the market's demand curve, $D(p)$.* Each firm produces a homogeneous product, so that there is a single price in this market.
5. *The dominant firm can predict how much output the competitive fringe will produce at any given price; that is, it knows the competitive fringe's supply curve, $S(p)$.*

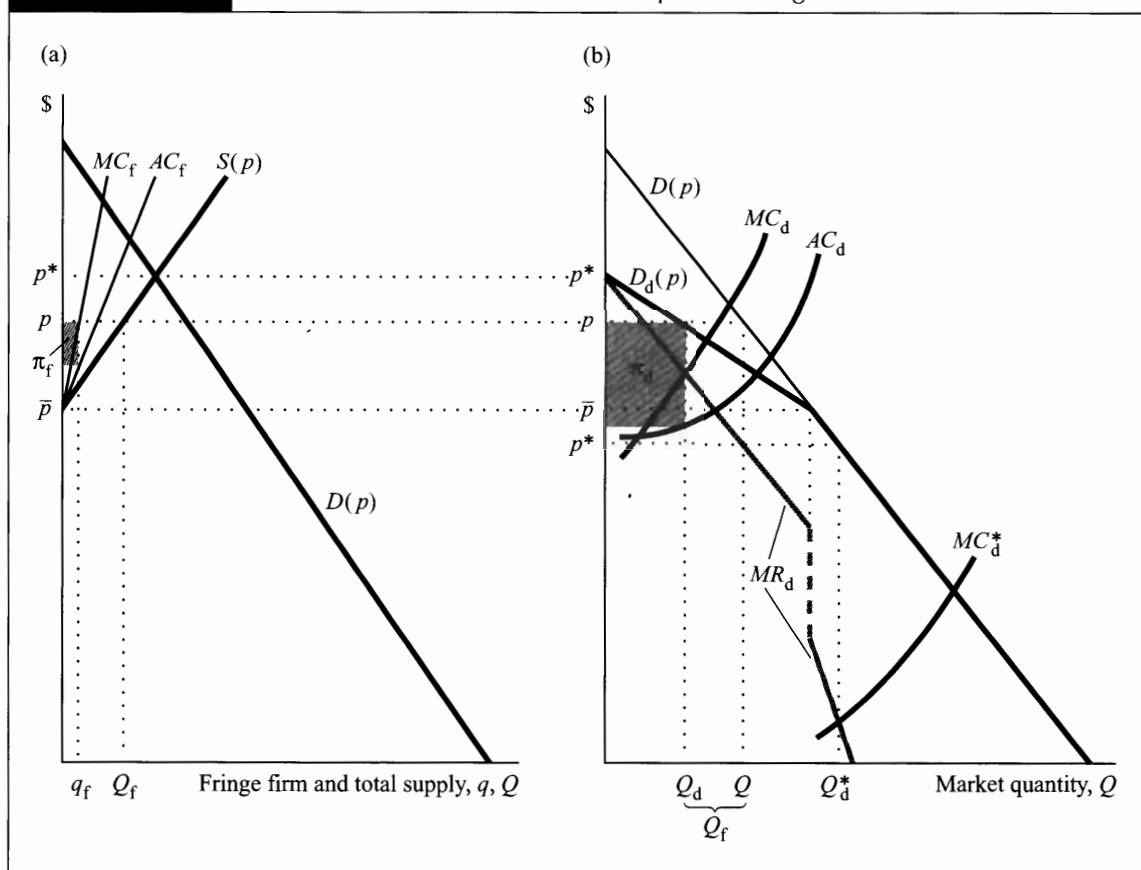
The first three assumptions determine that this market has a dominant firm facing a competitive fringe with no more than n firms. The last two assumptions ensure that the dominant firm knows enough to be able to set its output level optimally.

The Dominant Firm's Reasoning. Suppose you ran the dominant firm. How would you choose your output level? Given your firm's large size, you could drive up the market's price by restricting your output. Unfortunately for you, as your dominant firm lowers its output and price rises, the competitive fringe output increases because the fringe supply curve, $S(p)$, is increasing in p . As a result, market output falls less than you would like, and the market price does not rise as high as it would if your firm had a monopoly.

Thus, your dominant firm's problem is much more complex than that of a monopoly, which merely needs to consider the market demand curve (with its corresponding marginal revenue curve) and its marginal cost curve to determine its profit-maximizing output. Your dominant firm, in contrast, must consider not only those factors, but also how the competitive fringe responds to your actions.

To maximize your profits, you must take the competitive fringe's actions into account when setting your policy. A convenient way to calculate your optimal price level is to do the following thought experiment. For lack of an ability to stop them, let the fringe firms sell as much as they want at the market price: the price you set. Except at the very highest prices, the competitive fringe does not produce enough to meet all of the market's demand. Your dominant firm, then, is in a monopoly position with respect to this residual demand. Thus, you can determine your optimal output by a two-step procedure. First, determine your firm's residual demand curve; then, act like a monopoly with respect to the residual demand. This two-step procedure can be illustrated with the use of graphs.

A Graphic Analysis of Dominant-Firm Behavior. The first step is to determine the long-run residual demand curve facing the dominant firm. Figure 4.6 shows two graphs: (a) one for a representative competitive-fringe firm and for the entire competitive fringe, and (b) one for the dominant firm.

FIGURE 4.6**The Dominant Firm and the Competitive Fringe**

The graph on the left, Figure 4.6a, shows the market demand curve, $D(p)$, and the supply curve of a typical, price-taking, competitive-fringe firm. The fringe firm's supply curve is its marginal cost curve above the minimum of its average cost curve \bar{p} . That is, the fringe firm's shutdown price is \bar{p} . Above \bar{p} , each fringe firm makes positive economic profits. At \bar{p} , each fringe firm makes zero profits and is indifferent between operating and shutting down.¹⁶ Below \bar{p} , each firm shuts down, and the dominant firm is a monopoly.

The competitive fringe's supply curve, $S(p)$, is the horizontal summation of the individual fringe firm's supply curves, as Figure 4.6 shows. That is, $S(p) = nq_f(p)$, where n is the number of firms and q_f is the output of a typical fringe firm.

¹⁶As drawn, each fringe firm produces essentially no output at \bar{p} . If the firms had the usual U-shaped average cost curves, however, they would produce a positive amount of output at that price.

The dominant firm's residual demand curve is the horizontal difference between the market demand curve and the competitive fringe's supply curve:

$$D_d(p) = D(p) - S(p).$$

In Figure 4.6b, the market demand curve (thin blue line) is above the residual demand curve (heavy blue line) at prices above \bar{p} and equal to it at prices below \bar{p} . That is, the fringe firms meet some or all of the market demand if price is above \bar{p} , but they drop out of the market and leave all of the demand to the dominant firm if price falls below \bar{p} . At p^* , the quantity that the fringe supplies equals the quantity that the market demands, so the dominant firm has no residual demand.

The dominant firm maximizes its profits by picking a price (or equivalent, an output level) so that its marginal cost equals its marginal revenue. The dominant firm's marginal revenue curve (MR_d) is derived from its residual demand curve and has two distinct sections. If the competitive fringe produces positive levels of output, the dominant firm's residual demand curve lies below (and is flatter than) the market demand curve. The dominant firm's marginal revenue curve, MR_d , in this region is flatter than the marginal revenue curve in the region where the dominant firm's residual demand curve and the market demand curve are coincident. There is a discrete jump between the two sections of the marginal revenue curve at the point where the residual demand curve and the market demand curve meet.

The dominant firm behaves as a monopoly would with respect to the residual demand; it sets its price (or output) so that its marginal cost equals marginal revenue. Because the marginal revenue curve has two sections, there are two possible types of equilibria; which one occurs depends on the dominant firm's cost curves.

We now consider two types of markets:

1. The dominant firm charges a high price, so that it makes economic profits and the fringe firms also make profits or break even.
2. The dominant firm sets a price so low that the fringe firms shut down to avoid making losses. The dominant firm is now a monopoly.

The Dominant Firm–Competitive Fringe Equilibrium

The first type of equilibrium occurs if the dominant firm's costs are not substantially less than those of the fringe firms.¹⁷ The dominant firm's marginal cost curve, MC_d , crosses the first downward-sloping segment of the marginal revenue curve, MR_d , in Figure 4.6b.

The dominant firm chooses to produce Q_d level of output at price p (the height of the residual demand curve at the output level Q_d). At the price level p , the difference between the market demand, Q , and the dominant firm's output, Q_d , is the competitive fringe's supply, Q_f (which is shown in Figures 4.6a and 4.6b). If the dominant

¹⁷A mathematical analysis of this case is presented at www.aw-bc.com/carlton_perloff "Dominant Firm and Competitive Fringe Model."

firm's costs are this high, it does not drive the competitive fringe out of business. Its own profits are maximized at a price so high that the fringe firms make positive profits.

In most markets, positive economic profits would attract new entrants. In this market, however, no new firms can enter (by assumption), so both the dominant firm and the competitive fringe firms can make positive profits forever. In Figure 4.6b, the dominant firm's profits are labeled π_d . The profits of a typical fringe firm are positive as well (because $p > \bar{p}$), and a typical fringe firm's profits are shown as π_f in Figure 4.6a. Because the dominant firm's average cost is lower than that of the fringe firms (minimum $AC_d < \bar{p}$), the dominant firm makes more profits per unit (average profits), and it also sells more units than an individual fringe firm, so it must make more total profits as well.

Thus, the dominant firm maximizes its profits by charging a price so high that it loses some of its market share to the competitive fringe. It does not make sense for the dominant firm to set its price so low that it drives the fringe out of business, even though that would increase the number of units of output the dominant firm could sell. After all, few good business people accept the argument, "I lose a little on every sale, but make up for it in volume."

The dominant firm makes lower profits than it would if it were a monopoly and the fringe did not exist. The fringe can only hurt the dominant firm and benefit consumers. For example, in 1993, NEC Corporation, which then controlled half of all personal computer sales in Japan, had to cut its prices roughly in half due to increased competition from U.S. fringe firms.

The Dominant Firm as Monopoly. Now, suppose that the dominant firm has extremely low costs compared to the fringe firms, so that its marginal cost curve is MC_d^* in Figure 4.6b. Notice that MC_d^* crosses MR_d in the lower part of its two downward-sloping sections. The dominant firm chooses to produce Q_d^* level of output at price p^* (the height of the residual demand curve at output level Q_d^*). Because p^* is below the fringe firms' shutdown point (\bar{p} = their minimum average cost), the fringe firms produce nothing ($Q_f^* = 0$). As a result, market output, Q^* , equals the dominant firm's output, Q_d^* .

The dominant firm sets a monopoly price, and no competitive-fringe firm enters. The dominant firm meets all the demand of the market, unchecked by the fringe, and is thus a monopoly. The reason it has a monopoly is that MC_d^* intersects MR_d along the segment of MR_d that is the same as the marginal revenue curve associated with the market demand curve. That is, the monopoly price is below \bar{p} , so no fringe firm wants to produce.

A Model with Free, Instantaneous Entry

If unlimited entry is possible, a dominant firm cannot set as high a price as it can if entry is limited or prevented. This section retains all the assumptions made in the preceding section except that now an unlimited number of competitive-fringe firms may enter the market. Firms enter if they can make positive profits.

In this situation, fringe firms cannot make profits in the long run; they either break even or are driven out of business. If identical fringe firms produce at all, the market

EXAMPLE 4.9*China Tobacco Monopoly to Become a Dominant Firm*

Established in 1982, the Chinese government's tobacco monopoly, the China National Tobacco Corporation, has been the most profitable corporation in the world, accounting for 12% of the Chinese government's revenues. It sells to China's 310 million smokers, a quarter of the world's smoking population, who consume 1,700 billion cigarettes a year—about 30% of global consumption.

By imposing a 230% tax rate on foreign cigarettes, and by imposing import quotas and restrictions (such as designating only a few sales outlets for imported cigarettes), the government limited legal foreign cigarette sales to less than 2% of total Chinese sales in the late 1990s. By 2003, their share was only 10%.

To appease the World Trade Organization (WTO), China has agreed to lift restrictions on the retail sale of imported cigarettes by January 2004, to reduce the tariff on cigarettes from the current 65% to 24%, and to phase out the tariff over the next two years.

Thus, the state's monopoly will be turned into a dominant firm. Government officials expect that the price of imported cigarettes will drop in half, and that they will gain a major share of the market.

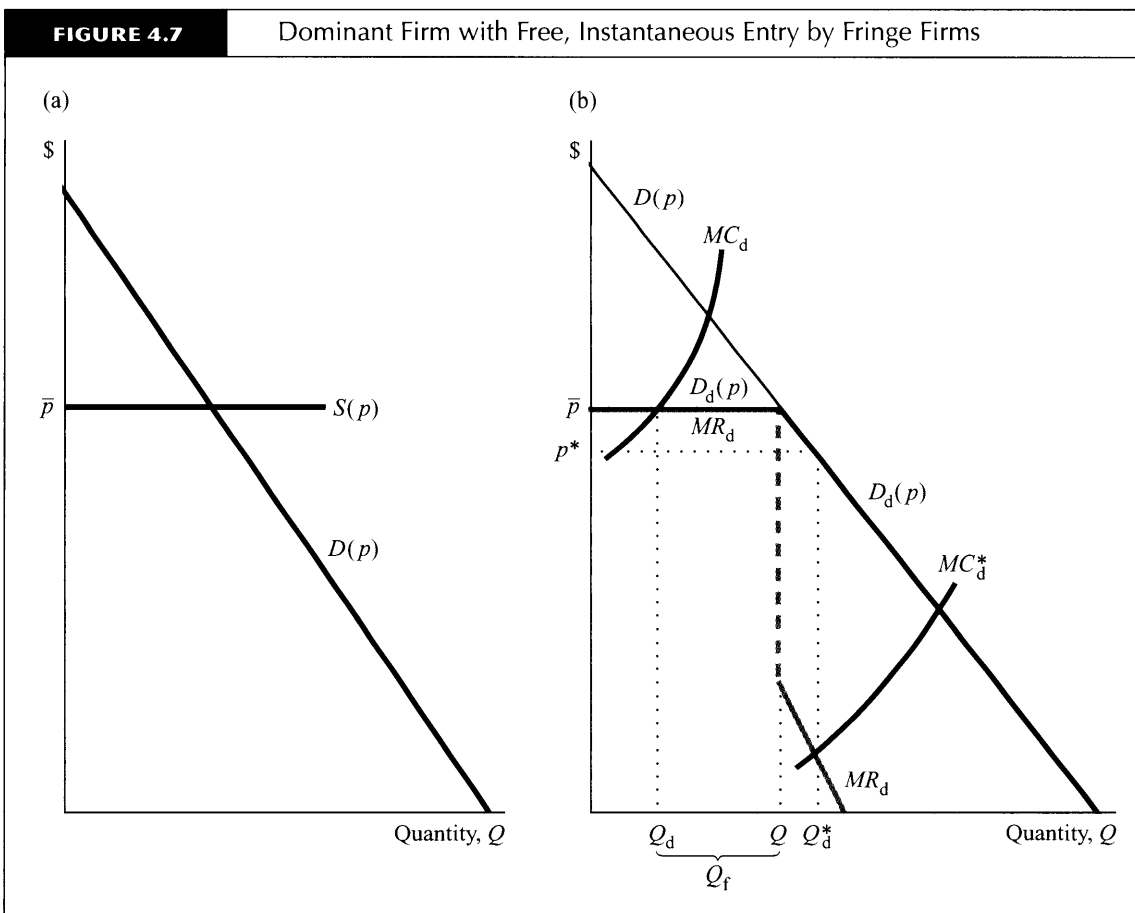
Sources: Glenn Collins, "U.S. Tobacco Industry Looks Longingly at the Chinese Market, but in Vain," *New York Times*, November 20, 1998:A10; "China to Lift Restrictions on Retail Sales of Imported Cigarettes Next Year," *AFX European Focus*, February 11, 2003; "Remove of Foreign Tobacco Retailing Licenses to Cut Prices by Half" (sic), *China News*, February 14, 2003:1; "Chinese Tobacco Industry Facing Mergers and Recapitalizations," *China Business Times*, February 17, 2003:1.

price ultimately can go no higher than a fringe firm's minimum average cost, so that fringe firms always just break even. After all, if they made positive profits, more firms would flood into the market and drive price down to the level where each earns zero economic profits. Because the dominant firm has lower costs than fringe firms, it makes positive profits, but its profits are lower than if entry did not occur.

Even with unlimited entry, the dominant firm can gain and hold indefinitely a large share of the market if it has some cost or other advantage (see Example 4.9). Another example is the Cheerleader Supply Co., which accounts for 60 percent of cheerleading uniforms and equipment sold in this country.¹⁸ This is an industry with easy entry, and yet one firm has the lion's share of the market, presumably because it has superior products, a superior sales force, lower costs, or has generated goodwill with buyers.

The competitive-fringe firms' cost curves are the same as before. As more and more firms enter (n rises), the slope of the competitive-fringe supply curve becomes flatter

¹⁸According to its chief executive officer, Lawrence Herkimer, in Peter Applebome, "The World's Oldest and Fattest Cheerleader," *San Francisco Chronicle*, January 12, 1984:24.



and flatter (it is n times the slope of a typical firm's supply, or MC , curve). As the number of firms grows large, the fringe's supply curve becomes essentially horizontal, as shown in Figure 4.7a. That is, as long as price is at least \bar{p} , the competitive fringe is capable of and is willing to supply any quantity that the market demands.

As shown in Figure 4.7b, the residual demand curve facing the dominant firm is horizontal at \bar{p} so the corresponding marginal revenue curve is also flat (remember that in a competitive market a firm faces a horizontal demand curve, and hence its marginal revenue curve is identical to its demand curve at the market price). Below \bar{p} the residual demand curve is the market demand, which slopes downward, so that the corresponding marginal revenue curve also slopes downward. Again, the marginal revenue curve corresponding to the residual demand curve jumps at the quantity where the kink in the residual demand curve occurs.

There are two possible equilibria. First, if the dominant firm's marginal cost is relatively high (MC_d in Figure 4.7b), so that it intersects the horizontal portion of the

MR_d curve, the price is \bar{p} , and the competitive fringe meets some of the market's demand. At this price, each fringe firm makes zero economic profits (because its average cost equals \bar{p}) and is indifferent between staying in business and leaving the market. How much is produced by the competitive fringe depends on the dominant firm's cost structure (that is, where MC_d intersects the horizontal marginal revenue curve), which determines the dominant firm's output, Q_d . Collectively, the fringe firms produce an output level $Q_f = Q - Q_d$, as Figure 4.7b shows.¹⁹ It is possible that $Q_f = 0$ even though the presence of the fringe constrains price to equal \bar{p} .

Thus, if fringe firms flood into a market whenever positive profits can be made, the dominant firm cannot charge a price above the minimum average cost of a fringe firm. Although a dominant firm can make positive profits, competitive-fringe firms just break even. If the dominant firm's price would be above \bar{p} in the absence of entry, consumers are better off if entry is possible because it results in lower prices.

The second type of equilibrium occurs if the dominant firm's marginal cost is lower (MC_d in Figure 4.7b), so that it hits the marginal revenue curve in the downward-sloping portion. Here, the price is so low that no fringe firm stays in the market when the dominant firm's costs are lower than the fringe firms' costs. This equilibrium (Q_d^*, p^*) is the same as discussed previously in the second no-entry equilibrium and is shown in Figures 4.6b and 4.7b. The dominant firm is a monopoly, and the potential supply of fringe firms is irrelevant.

SUMMARY

Monopoly or market power is the ability to price profitably above marginal cost. A single seller of a product, a monopoly, faces a downward-sloping demand curve and sets its price above marginal cost. As a result, less is purchased than if the market were perfectly competitive and society suffers a deadweight loss.

In some markets, however, there are benefits to monopoly. For example, the promise of future monopoly profits can spur a firm to develop new products or more efficient production techniques.

Not all firms that earn profits are monopolies, and not all monopolies earn profits. Just like a competitive firm, a monopoly can make either profits or losses in the short run. However, unlike a competitive firm, a monopoly can earn positive profits in the long run. A natural monopoly exists when it is efficient to have only one firm produce the market's output.

¹⁹Why don't fringe firms meet the entire demand at \bar{p} , instead of splitting it with the dominant firm? The answer is that the dominant firm has lower costs and can force some of the fringe firms out of the industry. Suppose that the dominant firm is producing its desired output of Q_d , and n fringe firms are producing $Q_f = Q - Q_d$. Now, if additional fringe firms enter this market, output exceeds market demand at \bar{p} . For the market to clear, the price must fall. Since the dominant firm is making positive profits, it stays in the industry. The fringe firms, however, start making losses (because they just break even at \bar{p}). Thus, some of the fringe firms must drop out of the industry until the price again rises to \bar{p} . Alternately stated, the dominant firm can always charge slightly below \bar{p} to sell as much as it wants.

Monopsony is monopoly on the buying side. A firm with monopsony power sets lower prices and employs fewer resources than would prevail under competition. Like monopoly, monopsony imposes an efficiency cost on society. Monopsony power can persist only when resources are specialized in the long run.

A low-cost dominant firm has market power even though it competes with other firms. A profit-maximizing dominant firm does not attempt to drive out fringe firms at all costs. Its behavior depends on how great its cost advantage over fringe firms is and on how easily other firms can enter. If a large number of price-taking firms can enter the market whenever a profit opportunity occurs, and if they can produce at costs not much above those of the dominant firm, the dominant firm is unable to charge prices substantially above the competitive price. Even if fringe firms do not enter a market, the threat of their entry may cause a monopoly (in the sense that it is the only firm in the market) to set a lower price than it would in the absence of the fringe.

PROBLEMS

1. If the demand curve is $Q(p) = 10 - p$ and the marginal cost is constant at 4, what is the profit-maximizing monopoly price and output? What is the price elasticity at the monopoly price and output?
2. Suppose that the supply of football players is elastic at the lowest salary levels paid. Would a monopsony of football players restrict output?
3. If the demand curve is $Q(p) = 5/p$, what is the elasticity of demand? What is total revenue when $p = \$1$ and when $p = \$30$? If production costs \$1 per unit, and the smallest production level is 1 unit, how much should the monopoly produce?
4. If the demand curve is $Q(p) = p^\epsilon$, what is the elasticity of demand? If marginal cost is \$1 and $\epsilon = -2$, what is the profit-maximizing price?
5. Suppose the demand curve for corn is $Q(p) = 10 - p$. Suppose that one firm owns all five units of corn in the world and has zero marginal cost. Does a monopoly sell less output than would be sold in a competitive market in which 100 firms each own 0.05 units?
6. Suppose the Environmental Protection Agency sets new requirements that raise the (fixed) costs of reporting compliance with pollution control rules (Pashigian 1984). How would this change affect (a) the market price, (b) the number of fringe firms, (c) total output, and (d) the dominant firm's share of the market? *Hint:* What does an increase in fixed costs do to the average cost curve of a fringe firm?
7. By showing the behavior of both a monopoly and a dominant firm in the same graph, show that monopoly profits are greater than the profit of a dominant firm in the no-entry equilibrium (MC_d). Show how much consumers benefit from buying from a dominant firm-competitive fringe rather than from a monopoly. *Hint:* A firm's variable costs are the area under its marginal cost curve up to the relevant output.
8. How would the no-entry model diagrams (Figure 4.6) change if fringe firms had the usual U-shaped average and marginal cost curves? Assume that because of a barrier to entry, there are only n fringe firms. Describe the types of possible equilibria.
9. Would a profit-maximizing dominant firm ever produce more than if it were a monopoly? *Hint:* Show the behavior of both a monopoly and a dominant firm (in the no-entry model) on the same graph and note where the marginal revenue curves cross.

10. What effect does a binding minimum wage have on a monopsony labor market?

Answers to odd-numbered problems are given at the back of the book.

SUGGESTED READINGS

Stigler (1965) provides a good, nontechnical introduction to the dominant firm-competitive fringe model. Fisher, McGowan, and Greenwood (1983)

is a very readable, controversial discussion of the important IBM antitrust case.

Industry Structure and Performance

Merely corroborative detail, intended to give artistic verisimilitude to an otherwise bald and unconvincing narrative. —W. S. Gilbert

Theories on competitive and noncompetitive markets hold that the less competition a firm faces, the greater its *market power*: the ability to set price profitably above marginal cost. Thus, market power (and hence price and profits) should be higher in industries with substantial entry barriers that reduce actual and potential competition. Economists conduct empirical investigations to test two of the implications of these theories:

1. How much market power do particular firms (industries) exercise?
2. What are the major factors that determine market power?

For many decades, economists have conducted *structure-conduct-performance* (SCP) studies that concentrate on the second question, which concerns the relationship between market performance and market structure. Market *performance* is the success of a market in producing benefits for consumers (for example, a market is performing well if prices are near the marginal cost of production). Market *structure* consists of those factors that determine the competitiveness of a market. Market structure affects market performance through the *conduct* or behavior of firms. Traditionally, SCP researchers presume that market power or performance can be measured relatively easily, and concentrate on the relationship between performance and structure.

In contrast, many economists now believe that readily available statistics often do *not* accurately reflect either market performance or structure. They rely

on new data and techniques to better measure the degree of market power, and its relationship to market performance.

This chapter starts with a summary of the theories on the major market structures based on Chapters 3 through 7. Then, it turns to SCP research and discusses the traditional SCP studies' measures of performance and analyses of the relationship between performance and structure. The main findings are that many industries appear to depart considerably from perfect competition, yet the degree of this departure apparently is not strongly related to industry concentration (the share of sales made by the largest firms in the industry), which presumably reflects the structure of the industry. Finally, the chapter examines modern studies of market power.

Theories of Price Markups and Profits

The relationship between price, p , and marginal cost, MC , and the existence and persistence of economic profits depend on the market structure (Table 8.1). In a competitive industry composed of identical firms with free entry, price equals short-run marginal cost; short-run profits, π_{SR} , are either positive or negative; and long-run profits, π_{LR} , are zero, where capital is charged at its rental price based on the competitive return (or normal return) that capital earns in a competitive industry. Even if firms are price takers (competitive), each firm's profit equals zero in the long run only if each firm has equal access to the same technology and inputs. If some firms have lower costs than others, their profits will not be eroded completely by entry. Free entry guarantees only that the profit of the least profitable firm to enter (the marginal firm) equals zero in the long run.

In monopoly or oligopoly, price exceeds marginal cost, profit in the short run is either positive or negative, and long-run profit is either zero or positive. In monopolistic competition, price is above marginal cost and entry drives long-run profit to zero.

Based on the relationships summarized in Table 8.1, two important conclusions can be drawn. First, testing whether long-run profits are positive is a test of free entry, not of (perfect) competition. Free entry guarantees that long-run profits equal zero, but

TABLE 8.1 Predictions Based on Market Structure

	$p - MC$	π_{SR}	π_{LR}
Competition	0	+ or –	0
Monopolistic competition	+	+ or –	0
Monopoly	+	+ or –	+ or 0
Oligopoly	+	+ or –	+ or 0

p = price, MC = marginal cost (short run), π_{SR} = short-run profits, and π_{LR} = long-run profits.

not that price equals marginal cost: Firms in a monopolistically competitive industry may earn zero profit even though price is above marginal cost. To determine whether price exceeds marginal cost, one must examine price data, not profit data. Second, short-run profits reveal very little about the degree of competition in an industry because, in all market structures, short-run profits can be either positive or negative.

Although Table 8.1 shows only four market structures, many more structures are possible. Moreover, for any given market structure, industries can differ substantially. For example, an oligopoly with four firms may set prices differently than one with only two firms. Generally, one would expect price-cost margins and profits to vary with the number of rivals and the size of barriers to entry. It is this generalization that provides the foundation for the SCP approach.

● Structure-Conduct-Performance

Edward S. Mason (1939, 1949) and his colleagues at Harvard introduced the structure-conduct-performance (SCP) approach, which revolutionized the study of industrial organization by introducing the use of inferences from microeconomic analysis. In the SCP paradigm, an industry's *performance*—its success in producing benefits for consumers—depends on the *conduct* or behavior of sellers and buyers, which depends on the structure of the market. The *structure* in turn depends on basic conditions such as technology and the demand for a product.

Because the nature of these connections is usually not explained in detail, many economists criticize the SCP approach for being descriptive rather than analytic. George J. Stigler (1968) and others argued that economists, rather than employ the SCP approach, should use price-theory models based on explicit, maximizing behavior by firms and governments. Others suggested replacing the SCP paradigm with analyses that emphasize game theory (von Neumann and Morgenstern 1944). We discuss modern approaches later in this chapter.

Most of the earliest SCP works were case studies of an individual industry (for example, Wallace 1937). The first empirical applications of the SCP theory were by Mason's colleagues and students, such as Joe S. Bain (1951, 1956). In contrast to the case studies, these studies made comparisons across industries.

A typical SCP study has two main stages. First, one obtains a measure of performance (through direct measurement rather than estimation) and several measures of industry structure. Second, the econometrician uses cross-industry observations to regress the performance measure on various measures of structure so as to explain the difference in market performance across industries. We first discuss the measurement of performance and structure variables and then examine the evidence relating performance to structure.

Measures of Market Performance

Measures of market performance try to provide an answer to our first key question as to whether market power is exercised in an industry. Two different measures that directly or indirectly reflect the profits or the relationship of price to costs are commonly used to gauge how close an industry's performance is to the competitive benchmark:

- The *rate of return*, which is based on profits earned per dollar of investment.
- The *price-cost margin*, which should be based on the difference between price and marginal cost, although, in practice, researchers often use some form of average cost in place of marginal cost.

A third measure, *Tobin's q*, is less commonly used. Tobin's *q* is the ratio of the market value of a firm to its value based on the replacement cost of its assets. (See www.aw-bc.com/carlton_perloff "Tobin's *q*" for more details.)

Rates of Return

A **rate of return** is a measure of how much is earned per dollar of investment. This section explains the relationship between economic profits and rates of return. The correct calculation of rates of return can be difficult, and sometimes compromises must be made that bias the final results. We discuss several different rate-of-return measures.

The Relationship Between Rates of Return and Economic Profit. The theories summarized in Table 8.1 make predictions about profit, and a rate of return is a measure of profit. The predictions in Table 8.1 refer to *economic* profit, which is revenue minus opportunity cost, not *accounting* profit (which is measured by accountants, using standard accounting principles). To test the predictions of Table 8.1, an economist's first step should be to adjust accounting profit to reflect economic profit before calculating the rate of return.

There are several important distinctions between economic and accounting profits. The main distinction concerns long-lived capital assets, like plant and equipment. Economic profit equals revenue minus labor, material, and an appropriate measure of capital cost. Measuring revenue, labor cost, and material cost is generally easy. The problem is measuring annual **capital cost**, which equals annual rental fees if all the capital assets were rented. The total rental fees equal the rental rate per unit times the number of units of capital. That is, the appropriate cost measure of capital is a *flow* (the price of renting capital per time period) and not a *stock* (the cost of capital, such as a machine, which lasts for many periods). If well-developed rental markets exist—for example, for used equipment—it is easy to calculate the relevant rental rate on capital and economic profits. When rental rates are not readily available, the economist must implicitly calculate a rental rate before calculating economic profit.

In the calculation of the implicit rental rate of capital used to determine long-run economic profits, capital assets should be valued at **replacement cost**, which is the long-run cost of buying a comparable-quality asset. If capital is valued at its replacement cost, then a low rate of return is a signal that no new capital should enter the industry. It does not mean that the firm should shut down or that it made an error in its past investment decisions. For example, a firm that bought machinery when it was cheap could earn a low rate of return based on the replacement cost and still have enjoyed a huge profit on its initial purchase. A high rate of return is a signal that new capital should enter the industry.

Researchers often divide economic profits by the value of the capital of the firm to obtain an earned rate of return on capital, which is a measure of profitability that controls for differences in capital across firms. There is a close relationship between economic profits, the earned rate of return on capital, and rental rates on capital. To develop this relationship requires an understanding of what a rental rate on capital really is: A rental rate must provide an owner of capital with a particular rate of return *after* depreciation has been deducted on the equipment.

Depreciation is the decline in economic value that results during the period the capital is used.¹ For example, if you rent your house for \$1,000 a year, and the wear and tear on the house is \$300 a year, then the depreciation is \$300, and your net annual rental after accounting for the depreciation is \$700. If the house is worth \$10,000 initially, then your rate of return is 7 percent, and the depreciation is 3 percent. What matters to the investor is the return after depreciation has been deducted. For that reason, a rental rate (per dollar of capital) can be expressed as an earned rate of return, r , plus a rate of depreciation, δ .

Your profit is

$$\pi = R - \text{labor cost} - \text{material cost} - \text{capital cost},$$

where R is revenue and capital costs are the rental rate of capital times the value of capital. The value of capital is $p_k K$, where p_k is the price of capital and K is the quantity of capital. If the rental rate is $(r + \delta)$, then profit is

$$\pi = R - \text{labor cost} - \text{material cost} - (r + \delta)p_k K. \quad (8.1)$$

The earned rate of return is that r such that economic profit is zero. Setting π equal to 0 and solving for r in Equation 8.1 yields

$$r = \frac{R - \text{labor cost} - \text{material cost} - \delta p_k K}{p_k K}. \quad (8.1')$$

Thus, the earned rate of return is net income divided by the value of assets, where *net income* is revenue minus labor cost minus material cost minus depreciation.²

The Relationship Between Rates of Return and Price. By how much would price or revenues have to fall in a highly profitable industry in order for that industry to earn a normal rate of return? To see how excess rates of return translate into price overcharges, suppose that a firm earns a rate of return r^* that is 5 percentage points higher than normal: $r^* = r + .05$. That is, the firm's invested capital earns excess rev-

¹An accountant's definition of depreciation may be based on a formula involving historical cost and age. This measure of depreciation is likely to differ from an economist's measure of depreciation, which is based on opportunity cost.

²Another rate-of-return measure is the internal rate of return, which is that interest rate such that the discounted present value of cash flows equals zero. The value of the internal rate of return is that it concisely summarizes the return earned by a project lasting several years. When profitability is changing over time, it may be misleading due to its aggregate nature. Because an internal rate of return depends primarily on the observed cash flows each year (except for the initial and terminal value of the firm), it frees the economist from having to calculate the value of capital each year.

enues of 5 percent times the value of its capital above what it would earn if it were in a competitive industry. If the firm's revenue is R^* , then its rate of return is

$$r^* = [R^* - \text{labor cost} - \text{material cost} - \delta p_k K] / p_k K = r + .05.$$

Let R be the revenue that would yield a normal rate of return, r . The amount by which revenue must decline to yield the normal return is $R - R^*$, all else constant. Using Equation 8.1' for r and the expression for r^* , we know that $r - r^* = -.05 = (R - R^*) / (p_k K)$. Multiplying both sides by $p_k K$, we find that $R - R^* = -.05 p_k K$. Thus, to get the normal rate of return, revenue would have to fall by 5 percent of the value of capital.

In many manufacturing industries, the ratio of the value of capital to the value of revenue is roughly 1. In such industries, revenues must fall by 5 percent in order for the industry to earn a normal return. Alternatively, all else constant, price needs to fall by 5 percent. Therefore, if a firm is earning a real rate of return 5 percent higher than the normal rate of return (which was roughly between 5 percent and 10 percent over the period 1948–1976), the competitive price is roughly 95 percent ($= 1 - .05$) of its current value. That is, industries that earn a rate of return 1.5 times higher than the return earned by competitive industries (say, 15 percent instead of 10 percent) have prices that are only 5 percent above those that generate a normal return. This price overcharge is the same as would occur if a monopoly faced an elasticity of -21 . In other words, even large differences in rates of return on capital between concentrated and unconcentrated industries do not necessarily imply that prices in concentrated industries are much above the competitive level. In industries with a low ratio of capital to revenue, even large excess returns can translate into tiny price overcharges.

Pitfalls in Calculating Rates of Return. There are eight major problems in calculating rates of return correctly (see Fisher and McGowan 1983). First, capital is usually not valued appropriately because accounting definitions are used instead of the economic definitions. An economist measures the annual capital cost flow as the annual rental fee if all the capital assets were rented.³ In contrast, the accounting value of capital, or *book value*, is based on the historical cost of the capital combined with accounting assumptions about depreciation. Capital should be valued at replacement cost (the long-run cost of replacing existing assets with comparable assets) to determine whether the rate of return is above the competitive level (in which case the firm or industry should expand) or below the competitive level (in which case the firm or industry should contract).⁴ Because historical cost is often very different from the actual replacement cost of the capital, using the book value of capital rather than the economic value can severely bias the measurement of rate of return.

³See www.aw-bc.com/carlton_perloff, Chapter 2, "Turning an Asset Price into a Rental Rate."

⁴In all but dying industries, the current value of capital depends on replacement cost. In dying industries, the value of capital is permanently less than replacement cost. The low value of capital is a signal that the industry should not invest in new facilities. In an expanding industry, the current value of capital can exceed replacement cost. The high value of capital is a signal that the industry should invest in new capital. The speed (and cost) of adjustment determines how long the current value can differ from replacement cost.

Second, depreciation is usually not measured properly. Accountants use several fixed formulas to measure the depreciation of an asset. One common formula, called *straight-line depreciation*, assumes that the asset's value declines in equal annual amounts over some fixed period (the *useful life* of the asset). For example, a machine that costs \$1,000 and is assigned a useful life of 10 years would incur \$100 of depreciation annually for its first 10 years of life. If it lasts more than 10 years, it incurs no additional depreciation. The fixed formula's predictions of the amount of depreciation may be unrelated to the asset's decline in economic value, which is the measure of its economic depreciation. As a result, the estimate of the rate of return may be biased. (See www.aw-bc.com/carlton_perloff "Accounting Bias in the Rate of Return.")

Third, valuing problems arise for advertising and research and development (R&D) for the same reason as for capital: All have lasting impacts on either a firm's demand or its costs. The money a firm spends on advertising this year may generate benefits next year, just as a plant built this year provides a benefit next year. If consumers forget about an advertisement's message slowly over time, the advertisement's effect on demand may last for several years. If a firm *expensed* (initially deducted its entire cost of) annual advertising expenditures and then made no deductions in subsequent years, its earned rate of return would be misleadingly low in the initial year and too high in later years. A better approach is to calculate the advertising cost based on the interest rate and the annual decline in the economic value of the advertising. Unfortunately, it is difficult to determine the correct rates of depreciation for advertising expenses.

Similar problems arise with R&D expenditures. Research and development can have a long-lasting impact. In addition, because R&D is risky, we need to be careful in interpreting rates of return. For example, suppose that a firm's research to discover new products is successful one time in ten. If the firm's expected profit is zero, then the profit on the successful product must be high enough to offset the losses on the nine failures. It is misleading to conclude that there are excessively high profits based on an examination of the profit of the one successful product.

Fourth, proper adjustment must be made for inflation. The earned rate of return can be calculated as either a *real* rate of return (a rate of return adjusted to eliminate the effects of inflation) or as a *nominal* rate (which includes the effects of inflation). One should be careful to compare rates that are either all real or all nominal.

If one is using a real rate, income in the numerator of the rate of return should not include the price appreciation on assets from inflation—it should only include the gain in the value of assets beyond that due to general price inflation. For example, if capital is initially worth \$100, annual income (before depreciation) is \$20, and the annual depreciation rate is 10 percent (so depreciation is \$10), then the earned rate of return is 10 percent $[(20 - 10)/100]$. If inflation was 20 percent during the year, the value of the capital at the end of the year equals \$90 ($\$100 - 10$ percent depreciation) times 1.2 (to adjust for inflation), or \$108. The firm has incurred a "gain" of \$18 on its capital, but it is illusory; it does not represent an increase in purchasing power because all prices have risen as a result of the inflation.

Fifth, monopoly profits may be inappropriately included in the calculated rate of return. This problem stems from using book value in the calculation, because book value sometimes includes *capitalized* (the present value of future) monopoly profit.

Suppose that the monopoly earns excess annual economic profits of \$100 above the competitive rate of return and the annual interest rate is 10 percent. The owner of the monopoly sells the firm (and its future stream of monopoly profits) for \$1,000 more than the replacement cost of its assets. The owner willingly sells the firm because that extra \$1,000 will earn \$100, or 10 percent, a year in a bank. The new owner makes only a competitive rate of return because the monopoly profit per year is exactly offset by the forgone interest payments from the extra \$1,000. The extra \$1,000 paid for the monopoly is the capitalized value of the monopoly profit, *not* the replacement cost to society of replacing the monopoly's capital. Thus, if the reported value of capital inappropriately includes capitalized monopoly profit, the calculated rate of return is misleadingly low if one wants to determine whether an industry is restricting output and is thereby earning an above-normal rate of return.

Sixth, the before-tax rates of return may be calculated instead of the appropriate after-tax rates of return. Corporations pay taxes to the government, and only what is left is of interest to individual investors. That is, after-tax rates of return govern entry and exit decisions. Competition among investors causes after-tax rates of return to be equated on different assets. If assets are taxed at different rates, the before-tax rate of return could vary widely even if all markets are competitive. For that reason, we should use after-tax rates of return and after-tax measures of profit, especially when comparisons are made across industries that are subject to different tax rates.

Seventh, rates of return may not be properly adjusted for risk. To determine whether a firm is earning an excess rate of return, the proper comparison is between the rate of return actually earned and the competitive **risk-adjusted rate of return**, which is the rate of return earned by competitive firms engaged in projects with the same level of risk as that of the firm under analysis. Investors dislike risk and must be compensated for bearing it: The greater the risk, the higher the expected rate of return.⁵

Eighth, some rates of return do not take debt into account properly. Researchers often use the rate of return to the stockholders as a measure of the firm's profitability. If a firm issues debt in addition to equity, both debtholders and equity holders (stockholders) have claims on the firm's income (Chapter 2). Because the assets of the firm are paid for by both debtholders and stockholders, the rate of return on the firm's assets equals a weighted average of the rate of return to the debtholders and the stockholders. The rate of return to debtholders is typically lower than the rate of return to stockholders, because debt is less risky than stock and debtholders get paid before stockholders when a firm is in financial distress. The return to stockholders increases with debt because the

⁵One commonly used approach to adjusting for risk is based on the Capital Asset Pricing Model. According to this model, the expected return on an asset equals the rate of return on risk-free investments (U.S. government Treasury bills are an example of a relatively risk-free investment) plus a number (called *beta*, the Greek letter β) times the difference between the market return (for example, return on the portfolio of all stocks) and the risk-free rate (Brealey and Myers 2003, Ch. 8). Beta reflects how closely the returns on one asset move with the returns on all other assets (the general economy). Risks that are related to movements in the general economy must generate higher returns than the riskless rate of return in order to attract investors.

income received by stockholders in a *highly leveraged* firm (one with a high ratio of debt to equity) is risky, so stockholders in such firms demand high rates of return.⁶

Therefore, it is improper to compare the rates of return to stockholders in two firms in order to measure differences in the degree of competition if the two firms have very different ratios of debt to equity. The debt/equity ratio has nothing to do with whether the firm is earning excess rates of return on its *assets*. Differences among firms in their rates of return to stockholders could reflect differences in competition facing firms or differences in their debt/equity ratios. Even though the rate of return calculated by dividing net income by assets differs from the rate of return from dividing income to stockholders by the value of stockholders' equity, they tend to be highly correlated (Liebowitz 1982b).

Comparing Rates of Return. To judge a rate of return, one must compare it to alternative rates of return. For example, if a firm has 100 units of capital each worth \$10, revenues of \$110, combined labor and material costs of \$10, and capital depreciation of 2 percent per year, then its earned rate of return is 8 percent per year: $(110 - 10 - 20)/1,000$. If investments in competitive industries yield a 5 percent rate of return, the firm is earning an *excess* rate of return.

There is an equivalent way to reach the same conclusion. If the rental rates on capital were based on the competitive rate of return of 5 percent, then the rental rate would equal 7 percent (5 percent plus depreciation of 2 percent). Calculating economic profit as revenue minus labor cost, material cost, and capital (rental) cost yields a *positive* economic profit of \$30 $(110 - 10 - [.07 \times 1,000] = 30)$. Thus, earning *positive* economic profit and earning *excess* rates of return (above the competitive or normal level) are equivalent ways of expressing the same idea. Excess economic profit exists if the earned rate of return exceeds the competitive rate.

Fraumeni and Jorgenson (1980) calculated the after-tax economic rate of return for a large sample of American industries over the period 1948–1976. In their calculations, they were careful to avoid many of the pitfalls described above. They found that over this period, the median manufacturing industry earned a nominal (unadjusted for inflation) rate of return of approximately 11 percent (Table 8.2). Over this same pe-

⁶Suppose that a firm initially has no debt and finances an investment with \$1,000 raised through sale of stock. Next year, the investment returns the \$1,000 plus either \$80 or \$200 with equal probability, so that stockholders' rate of return is either 8 percent or 20 percent, for an average return of 14 percent. Suppose, instead, that the firm raises the \$1,000 for the investment by issuing debt of \$500 that pays 10 percent interest and selling stock worth \$500. Debtholders must receive payment of interest before stockholders receive any income. Therefore, whether the firm earns \$80 or \$200, debtholders receive \$500 plus \$50 of interest. Stockholders receive \$500 plus either \$30 or \$150, so that the total amount paid to both debtholders and stockholders is \$1,000 plus either \$80 or \$200. Stockholders therefore earn either 6 percent $(= 30/500)$ or 30 percent $(= 150/500)$, for an average return of 18 percent, while debtholders earn 10 percent. Stockholders now earn a higher average rate of return and face a wider range of outcomes, even though the income potential of the firm is unchanged.

TABLE 8.2**Average Annual Returns, 1948–1976**

Industry	Nominal Rate of Return	Own Rate of Return*	Nominal Rate of Return on Stockholders' Equity
Agriculture	.07	.04	
Crude petroleum	.12	.08	
Food	.10	.07	.10
Tobacco	.14	.11	.13
Textiles	.09	.06	.08
Chemicals	.13	.10	.14
Motor vehicles	.29	.25	.15
All manufacturing - median industry	.11	.08	.11
Railroads	.07	.03	
Telephone and telegraph	.15	.11	
Retail trade	.10	.07	

*The own rate of return subtracts from income the effects of increases in the price of capital for each industry. If the price of capital changes only with inflation, the own rate of return is a real (inflation-adjusted) rate of return.

Sources: Fraumeni and Jorgenson (1980); Federal Trade Commission, *Quarterly Financial Reports*, 1948–1976.

riod, the average rate paid on three-month U.S. government Treasury bills was roughly 3.6 percent, so the rate of return in manufacturing significantly exceeded the rate of return on Treasury bills, possibly to compensate for the increased risk.

Studies that calculate rates of return often differ in their methodologies and, because of data constraints, are commonly forced to calculate something other than economic rates of return. Nevertheless, they can still be valuable in investigating whether the rate of return in one industry is higher than that in another, as long as the biases in the calculated rates of return are similar across different industries. It is dangerous, however, to compare the absolute levels of rates of return from one study with the absolute levels of rates of return from another study if the studies follow different methodologies for calculating rates of return.

To illustrate the differences that can arise when different concepts are used to calculate rates of return, the last column of Table 8.2 presents the returns on the book value of stockholders' equity (the difference in the book values of assets and liabilities) published by the Federal Trade Commission (FTC). These rates of return are calculated as after-tax corporate income (which deducts interest payments on debt) divided by the stockholders' equity in the company. Table 8.2 shows that different methodologies can lead to different rates of return. For example, the nominal rate of return in motor vehicles is about 29 percent according to Fraumeni and Jorgenson (1980), but is 15 percent according to the FTC. Nonetheless, the relative rates of return between industries follow the same pattern using both methodologies. For example, tobacco earns a higher rate of return than textiles according to Fraumeni and Jorgenson as well as the FTC.

Price-Cost Margins

To avoid the problems associated with calculating rates of return, many economists use a different measure of performance, the Lerner Index or *price-cost margin*, $(p - MC)/p$, which is the difference between price, p , and marginal cost, MC , as a fraction of the price. The predictions in the first column of Table 8.1 about the relationship of price to marginal cost are stated in terms of the price-cost margin. Because the correlation between accounting rates of return and the price-cost margin can be relatively low (Liebowitz 1982b), it makes a difference which of these two performance measures is used.

The price-cost margin (Chapter 4) for a profit-maximizing firm equals the negative of the reciprocal of the elasticity of demand, ϵ , facing the firm:

$$\frac{p - MC}{p} = -\frac{1}{\epsilon}. \quad (8.2)$$

A competitive firm sets $p = MC$ because its residual demand price elasticity is negative infinity (it faces a horizontal demand curve).

Unfortunately, because a marginal cost measure is rarely available, many researchers use the price-average variable cost margin instead of the appropriate price-marginal cost margin.⁷ Their approximation to the price-average variable cost margin is typically calculated as sales (revenues) minus payroll minus material cost divided by sales. That is, they tend to ignore capital, research and development, and advertising costs.⁸

This approach may lead to serious biases. Suppose that marginal cost is

$$MC = v + (r + \delta) \frac{p_k K}{Q}, \quad (8.3)$$

where r is the competitive rate of return, δ is the depreciation rate, and the cost of the labor and materials needed to produce 1 unit of output, Q , is v . Equation 8.3 describes a technology that requires K/Q units of capital (at a cost of p_k per unit of capital) to produce 1 unit of output. Using v in place of marginal cost can lead to serious bias, however, as can be seen by substituting MC from Equation 8.3 into Equation 8.2 to obtain

$$\frac{p - v}{p} = -\frac{1}{\epsilon} + (r + \delta) \frac{p_k K}{pQ}. \quad (8.4)$$

Thus, $(p - v)/p$ differs from the correct measure $(p - MC)/p = -1/\epsilon$ by the last term in Equation 8.4, $(r + \delta)p_k K/(pQ)$, which is the rental value of capital divided by the value of output.

⁷A few studies (Keeler 1983, Friedlaender and Spady 1980) estimate marginal cost based on cost functions.

⁸See Fisher (1987) for a critique of the typical price-cost margin. An even more serious error that is sometimes made is to use average total cost.

Measures of Market Structure

To examine how performance varies with structure, we also need measures of market structure. A variety of measures are used, all of which are thought to have some relation to the degree of competitiveness in an industry. We now describe some of the common measures of market structure.

Industry Concentration. In most SCP studies, industry concentration is the structural variable that is emphasized. Industry concentration is typically measured as a function of the market shares of some or all of the firms in a market.

By far, the most common variable used to measure the market structure of an industry is the four-firm concentration ratio, C_4 , which is the share of industry sales accounted for by the four largest firms. It is, of course, arbitrary to focus attention on the top four firms in defining concentration ratios. Other concentration measures are used as well. For example, the U.S. government also has published eight-firm concentration ratios, C_8 .

Alternatively, one could use a *function* of all the individual firms' market shares to measure concentration. The most commonly used function is the **Herfindahl-Hirschman Index**, HHI, which equals the sum of the squared market shares of each firm in the industry. For example, if an industry has three firms with market shares of 50, 30, and 20 percent, the HHI equals 3,800 ($= 2,500 + 900 + 400$). More attention has been paid to the HHI since the early 1980s, when the Department of Justice and Federal Trade Commission started using it to evaluate mergers. The government publishes HHI statistics by industry.

Typically, empirical studies produce similar results for both the HHI and a four-firm concentration index. It has been shown theoretically (Appendix 8A) that the HHI is the appropriate index of concentration to explain prices if firms behave according to the Cournot model.

Rather than aggregating information about the relative sizes of firms into a single measure, one could examine the effects of the market shares of the first, second, third, fourth, and smaller firms on industry performance. For example, one could determine whether increases in the market share of the second firm raise prices by as much as increases in the share of the leading firm. Using this approach, Kwoka (1979) showed that markets with three (relatively equal-size) firms are much more competitive than those with only two firms.

Table 8.3 shows three concentration measures— C_4 , C_8 , and HHI—for several manufacturing industries. Aside from concentration in individual industries, one can examine concentration in manufacturing in general. The 1997 Census of Manufactures reports concentration ratios for 470 manufacturing industries. In 1997, the concentration ratio of the four largest firms was below 40 percent in more than half of the industries, between 41 and 70 percent in about one-third of the industries, and over 70 percent in about one-tenth of the industries, based on value of shipments.

There are now more industries with low four-firm concentration ratios and fewer with high four-firm concentration ratios than in 1935. In 1935, about 47 percent of industries had a four-firm concentration ratio below 40 percent, and about 16 percent of industries had ratios above 70 percent. Since World War II, however, the distribution of concentration ratios in manufacturing has not changed much. Comparisons

TABLE 8.3 1997 Concentration Ratios in Selected Manufacturing Industries

Product Grouping	C4	C8	HHI*
Meat products	35	48	393
Breakfast cereal	83	94	2,446
Distilleries	60	77	1,076
Cigarettes	99	NR	NR
Men's and boy's suits and coats	42	56	846
Sawmills	15	20	87
Folding paperboard boxes	25	38	246
Book printing	32	45	364
Petroleum refining	29	49	422
Tires and inner tubes	68	86	1,518
Blast iron and steel mills	33	53	445
Household refrigerators and freezers	82	97	2,025
Motor vehicles and car bodies	87	94	NR
Computers	40	68	658

*Herfindahl-Hirschman Index for the 50 largest companies. *NR* indicates that the index is not reported.

Source: *Census of Manufactures: Concentration Ratios in Manufacturing* (2001, Table 2).

based on value of shipments, and not on the number of industries, produce similar conclusions.

Table 8.4 shows that there has not been a trend toward increasing aggregate concentration in the manufacturing sector based on *value added* (revenue minus the cost of fuel, power, and raw materials) accounted for by the largest firms. The table shows that aggregate domestic concentration has increased since 1947, but remained relatively constant between 1967 and 1992 and fell slightly in 1997. Moreover, these domestic concentration statistics overstate concentration because they ignore imports, which have grown in importance.

Most of what we know about concentration ratios concerns manufacturing industries, which comprised only about 14 percent of the GDP in 2001.⁹ What about concentration in the other sectors of the economy? Unfortunately, data on concentration ratios are not readily available for most individual industries outside of manufacturing. It is generally believed that ease of entry keeps most of agriculture, services, retailing and wholesale trade, and parts of manufacturing and finance, real estate, and insurance relatively unconcentrated.

Unfortunately, concentration measures have two serious problems. First, many factors influence seller concentration measures. For example, profitability may affect the degree of concentration in an industry by affecting entry. One of the key questions

⁹Table B-12, Economic Report of the President, 2003.

TABLE 8.4		Percent Aggregate Concentration in the Manufacturing Sector (measured by value added)								
Top Firms	1947	1954	1963	1967	1972	1977	1982	1987	1992	1997
50 largest	17	23	23	25	25	24	24	25	24	21
100 largest	23	30	30	33	33	33	33	33	32	29
200 largest	30	37	38	41	42	43	43	43	42	38

Sources: 1982, 1987, 1992, and 1997, *Census of Manufactures: Concentration Ratios in Manufacturing*, Table 1.

posed in the introduction concerns whether a less competitive market structure “causes” higher profits. A test of this hypothesis is meaningful only if structure affects profits, but not vice versa. That is, this theory should be tested using *exogenous* measures of structure, where exogenous means that the structure is determined before profitability and that profitability does not affect structure.¹⁰

Most commonly used measures of market structure are not exogenous. They depend on the profitability of the industry. For example, suppose that we use the number of firms as a measure of the structure of an industry, arguing that industries with more firms are more competitive. However, entry occurs in extraordinarily profitable industries if there are no barriers to such entry. Although, in the short run, an inherently competitive industry may have a small number of firms, in the long run, many additional firms enter if profits are high.

An exogenous barrier to entry is a better measure of structure than the number of firms. For example, if a government historically prevented entry in a few industries, those industries with the barrier should have higher profits, but the higher profits do not induce additional entry.

Most SCP studies have ignored the problem with obtaining exogenous measures of market structure. In particular, the commonly used concentration measures, such as C4, are definitely *not* exogenous measures of market structure.

The second serious problem is that many concentration measures are biased because of improper market definitions. The relevant *economic market* for a product includes all products that significantly constrain the price of that product (see Chapter 19). In order for industry concentration to be a meaningful predictor of performance, the industry must comprise a relevant economic market. Otherwise, concentration in an industry has no implication for pricing.

For example, the concentration ratio for an industry whose products compete closely with those of another industry may understate the amount of competition. If plastic bottles compete with glass bottles, the concentration ratio in the glass-bottle industry may reveal very little about market power in that industry. The relevant concentration measure should include firms in both industries. Similarly, firms classified in

¹⁰If measures of structure are determined by profitability, the measures are said to be *endogenously* determined. Failure to use exogenous measures of structure leads to what statisticians call the “simultaneous equations estimation problem.”

one industry that can modify their equipment and easily produce products in another industry are potential suppliers that influence current pricing, but are not reflected in the relevant four-firm concentration ratio.¹¹

Unfortunately, concentration ratios are published by the government for specific industries and products, and the definitions used do not necessarily coincide with relevant economic markets. Concentration measures are often based on aggregate national statistics. If the geographic extent of the market is local because transport costs are very high, national concentration statistics may misleadingly indicate that markets are less concentrated than is true. Some researchers use distance shipped to identify markets in which the use of national data is misleading: If the distance shipped is short, the concentration in the local market may be much different from the national market concentration.

Similarly, concentration measures are often biased because they ignore imports and exports. For example, the 1997 four-firm concentration ratio for U.S. automobiles was 80 percent. This figure indicates a very concentrated industry; however, it ignores the imports of British, Japanese, and German cars, which were over 23 percent of total 1997 sales in the United States. The use of improper concentration measures, of course, may bias the estimates of the relationship between performance and concentration.

Just as seller concentration can lead to higher prices, buyer concentration can lead to lower prices. When buyers are large and powerful, their concentration can offset the power of sellers. For that reason, several researchers include buyer concentration as a market structure variable explaining industry performance. The same type of market definition problems can affect this measure. However, this measure is more likely to be exogenous than is seller concentration.

Barriers to Entry. Probably the most important structural factor determining industry performance is the ability of firms to enter the industry (Chapter 4). In industries with significant long-run entry barriers, prices can remain elevated above competitive levels.

Commonly used proxies for entry barriers include minimum efficient firm size, advertising intensity, and capital intensity, as well as subjective estimates of the difficulty of entering specific industries. Chapter 3 makes a distinction between a long-run barrier to entry and the speed with which entry can occur. Most empirical studies do not distinguish these two concepts, and so any measure they use for entry barrier typically reflects both concepts.

Fraumeni and Jorgenson (1980) show that differences in rates of return across industries persist for many years. If there are no long-run barriers to entry or exit, rates of return across industries should converge. Their results indicate that there are long-run barriers, or that the rate of entry and exit is very slow so that convergence in rates of return is slow across industries, or that there are persistent differences across industries in the levels of risk that are reflected in rates of return.

¹¹If the producers of some Product B could profitably switch production to Product A (Product B is a supply substitute for Product A), then the producers of Product B should also be considered in the market for Product A.

Again, many of the proxies to barriers to entry, such as advertising intensity, are not exogenous. Others, such as subjective measures, have substantial measurement bias.

Unionization. If an industry is highly unionized, the union may be able to capture the industry profits by extracting them through higher wages. Moreover, the higher wages would drive prices up. Therefore, unionization may raise prices to final consumers even though profits of the firms in the industry are not excessive. It is also possible that unions could raise wages and prices and also raise profits to the industry. By making it costly to expand the labor force, unions can prevent industry competition from expanding output and driving profits down. Unionization may not be exogenous if unions are more likely to organize profitable industries.

The Relationship of Structure to Performance

There are hundreds, if not thousands, of studies that attempt to relate market structure to each of the three major measures of market performance. This section first discusses the key empirical findings for each of the performance measures based on U.S. data.¹² Then, SCP studies based on data from other countries and on data for individual industries are examined. Finally, the section summarizes the major critiques of the results and their interpretation.

Rates of Return and Industry Structure. Joe Bain deserves credit for pioneering work that led to the voluminous literature on the relationship between rates of return and industry structure. Bain (1951) investigated 42 industries and separated them into two groups: those with an eight-firm concentration ratio in excess of 70 percent and those with an eight-firm concentration ratio below 70 percent. The rate of return (calculated roughly as income divided by the book value of stockholders' equity) for the more concentrated industries was 11.8 percent compared to 7.5 percent for less concentrated industries.

Bain (1956) classified industries by his subjective estimate as to the extent of barriers to entry. His hypothesis was that profits should be higher in industries with high concentration and high barriers to entry. The evidence that Bain presented is consistent with his hypothesis.

Brozen (1971) criticized Bain's findings for two reasons. First, as Bain recognized, the industries that Bain studied could be in disequilibrium. Brozen showed that the industries Bain identified as highly profitable suffered a subsequent decline in their profits, while the industries of lower profitability enjoyed a subsequent increase in profits. In fact, for the 42 industries of Bain's initial 1951 study, the profit difference of 4.3 percent that he found between the highly concentrated and less concentrated groups diminished to only 1.1 percent by the mid-1950s (Brozen 1971). Second, Brozen pointed out that Bain's use, in some of his work, of the profit rates of the leading firms, rather than the profit rate of the industry, could have skewed his results.

¹²See also www.aw-bc.com/carlton_perloff "Tobin's q " for a discussion of research using Tobin's q .

TABLE 8.5**Average Profit Rates (selected industries)**

Eight-Firm Concentration Ratio over 70 Percent		Eight-Firm Concentration Ratio below 70 Percent	
Industry	Profit Rate (%)	Industry	Profit Rate (%)
Auto	15.5	Shoes	9.6
Cigarettes	11.6	Beer	10.9
Ethical drugs	17.9	Bituminous coal	8.8
Liquor	9.0	Canned fruits and vegetables	7.7
Steel	9.0	Average for all industries studied	9.0
Average for all industries studied	13.3		

Source: Mann (1966, 299).

Using 1950–60 data, Mann (1966) reproduced many of Bain's original findings (Table 8.5). Using the same 70 percent concentration ratio criterion as Bain used to divide his sample into two groups, Mann found that the rate of return for the more highly concentrated group was 13.3 percent compared to 9.0 percent for the less concentrated group.

Mann also investigated the relationship between profit and his own subjective estimates of barriers to entry. He found that industries with “very high” barriers to entry enjoy higher profits than those with “substantial” barriers, which in turn earn higher profits than those with “moderate to low” barriers. He confirmed Bain's predictions and earlier findings that concentrated industries with very high barriers to entry have higher average profit rates than concentrated industries that do not have very high barriers to entry.

There have been many econometric estimates of the relation between rates of return, concentration, and a variety of other variables, such as those measuring barriers to entry (see the surveys by Weiss 1974 and Schmalensee 1989). Econometric studies attempt to measure the effects of several variables on rates of return. Such an estimated relationship is called a *regression*. Regression studies provide not only an estimate of the effect of one variable on another but also a statistical measure of whether the estimated effect could be different from zero.

Based on his survey of many of these studies, Weiss (1974) concluded that there was a significant relationship between profit, concentration, and barriers to entry. Studies based on more recent data tend to find only a weak relationship or no relationship between the structural variables and rates of return. For example, Salinger (1984) found, at best, weak support for the hypothesis that minimum efficient scale in concentrated industries is related to rates of return.¹³ He found no statistical support that

¹³Large capital requirements do not constitute a long-run barrier to entry unless other conditions, such as imperfect capital markets or sunk costs, are present (see Chapter 3).

his other entry barrier proxy variables (such as advertising intensity) are related to rates of return.

Econometric studies linking profit to market structure often conclude that measured profitability is correlated with the advertising-to-sales ratio and with the ratio of research and development expenditures to sales. These studies also commonly find that high rates of return and industry growth are related.

Some researchers have studied how the speed of adjustment of capital (and hence profit) is related to concentration. Capital-output ratios appear to rise with concentration, though less so recently than in the past. (Table 8.6). The full explanation for the correlation between capital-output ratios and concentration is not known. One possible reason for this result is that the plant of minimally efficient scale (the smallest plant that can operate efficiently) is so large relative to industry size that when economies of scale are important, only a few of them can fit into the industry. However, for most industries, minimum efficient scale (Chapter 2) is a small fraction of total industry demand.

It is possible that the more capital-intensive, concentrated industries use relatively more specialized capital. If so, their rates of adjustment of output should be slower than those of less concentrated industries because it is usually more difficult to adjust specialized capital than it is to adjust less specialized capital. If highly concentrated industries adjust more slowly than unconcentrated industries, that explains why high (or low) profits take longer to fall back to (rise to) the industry average in these industries (Stigler 1963, Connolly and Schwartz 1985, Mueller 1985). See Chapter 3 for studies of entry.

TABLE 8.6 Capital-Output Ratios and Concentration

Four-Firm Concentration Ratio	Average Capital/Output Ratio (Percent)	
	1963	1997
0–10	26.5	38.8
11–20	26.9	32.8
21–30	32.7	37.1
31–40	34.5	39.9
41–50	37.7	36.8
51–60	37.9	39.4
61–70	44.2	46.6
71–80	49.8	49.0
81–90	51.8	35.6
91–100	57.7	42.5

Source: 1963 series from Collins and Preston (1969, 272); 1997 series is based on authors' calculations using the 1997 *Census of Manufactures, Industry Series and Concentration Ratios in Manufacturing*. The numbers in Table 8.6 are based on gross book value of capital and, because of data unavailability, do not exclude depreciation.

Similarly, if concentrated industries take a long time to react to demand changes, then, all else equal, good economic news should raise the value of a company in a concentrated industry more than the value of a company in an unconcentrated industry. Lustgarten and Thomadakis (1980) find that good economic news raises the stock market values of companies in concentrated industries much more than those in unconcentrated industries, and bad economic news lowers their values more.

Price-Cost Margins and Industry Structure. Following Collins and Preston (1969), many economists examine the relationship across industries between price-average variable cost margins based on Census data and various proxies for industry structure, such as the four-firm concentration ratio and the capital-output ratio. A typical regression based on data from 1958 (Domowitz, Hubbard, and Petersen 1986, 7) is

$$\frac{p - v}{p} = .16 + .10 C4 + .08 \frac{p_k K}{pQ} + \text{other variables},$$

(.01) (.02) (.02)

where $(p - v)/p$ is the price-average variable cost margin, v is a measure of average variable cost, $C4$ is the four-firm concentration ratio, and $p_k K/(pQ)$ is the ratio of the book value of capital to the value of output. The numbers in parentheses below each coefficient are standard errors, which are a measure of how precisely the coefficients are estimated.¹⁴ The $p_k K/(pQ)$ term is necessary because price-average variable cost margins are used (see Equation 8.4).

The sensitivity of price to increases in concentration can be derived from this equation. According to the equation, if the value of capital to output, $p_k K/(pQ)$, is 40 percent (the average value across industries), the concentration ratio of the top four firms, $C4$, is 50 percent, and if other variables are zero, the predicted price-average variable cost margin is .24 ($\approx .16 + [.10 \times .5] + [.08 \times .4]$), or $p = 1.3v$. That is, price is 30 percent above average variable cost.

If this industry's four-firm concentration ratio doubles from 50 percent to 100 percent, the price-average variable cost margin rises to .29 or $p = 1.4v$. That is, price rises to approximately 1.4 times average variable cost, which is an increase in the price of only about 7 percent. Thus, even very large increases in concentration may raise price by relatively modest amounts.

Domowitz, Hubbard, and Petersen (1986) found that, for the time period 1958–1981, the differential in the price-average variable cost margins between industries of high and low concentration fell substantially over time. When they estimated a price-average variable cost equation with more recent data, the coefficient associated

¹⁴A commonly used method to express the confidence one has in a coefficient is to construct an interval (called a "95 percent confidence interval") for a coefficient that includes all values within roughly two standard errors of the estimated coefficient. The 95 percent confidence interval for the coefficient on $C4$ covers .06 to .14.

with the concentration ratio is much lower than its value in 1958. That is, the already small effect of concentration on price in 1958 shrunk in later years. Further, in the later period, a statistical test of the hypothesis that the concentration measure does not affect the price-average variable cost margin could not be rejected. In general, they found that the relationship between price-cost margins and concentration is unstable, and, to the extent that any relationship exists, it is weak, especially in recent times.

Instead of using industry average variable cost Census data to study the relationship between the price-average variable cost margin and industry structure, other investigators, among them Kwoka and Ravenscraft (1985), used Federal Trade Commission (FTC) data to investigate price-average variable cost margins at the individual firm level.¹⁵ The studies using individual firm data showed that the link between higher concentration and higher price-cost margins is ambiguous. Some studies find that the link, if it exists at all, is very weak, whereas others discern no link at all. They also find that the presence of a large second or third firm greatly reduces the price-cost margin that can be earned. This discovery indicates that it is a mistake to use only four-firm concentration ratios to measure market structure.

Various studies report significant effects from other explanatory variables. Kwoka and Ravenscraft (1985) showed that industry growth has a significant and positive effect on price-average variable cost margins. Lustgarten (1975b) concluded that increased buyer concentration sometimes lowers price-cost margins. Comanor and Wilson (1967) reported that a higher advertising-sales ratio may raise the price-cost margin. Freeman (1983) showed that unions lower the price-cost margin.¹⁶

International Studies of Performance and Structure. Because international trade is more important in many other countries than it is in U.S. markets, the bias from ignoring imports and exports may be more substantial in studies based on data from those countries than on U.S. data. Concentration ratios based only on domestic concentration may not be economically meaningful as measures of market power. The relevant competition may well be from firms located outside a given country.

Nonetheless, despite differences across countries in sizes of domestic markets, domestic concentration ratios are correlated across countries (Pryor 1972). That is, an industry that is concentrated in the United States is also likely to be concentrated in the United Kingdom. However, the correlation is not perfect, as illustrated by Sutton (1989, 1998) for the U.S. and U.K. frozen food industries.

¹⁵The advantage of using the firm rather than the industry as the unit of observation is that the researcher can disentangle the effect of industry concentration on a firm's price-cost margin from the effect of the efficiency of that firm alone. For example, one firm's price-cost margin may be high either because the firm is particularly efficient (low cost relative to all other firms) or because all firms in the industry enjoy a high price (lack of competition in the industry). See Benston (1985) for a critique of studies that rely on the FTC data.

¹⁶Salinger (1984) and Ruback and Zimmerman (1984) also found that unionism has a significant negative effect on the profits of highly concentrated industries. Voos and Mishel (1986) showed that, although unions may depress the price-cost margin, the price is not significantly above the one that would prevail in the absence of a union.

Regardless of which country's data are used, most studies have difficulty detecting an economically and statistically significant effect of concentration on performance (Hart and Morgan 1977, Geroski 1981). However, Encoau and Geroski (1984) found that the United States, the United Kingdom, and Japan tend to have slow rates of price adjustment in their most concentrated sectors.¹⁷

Performance and Structure in Individual Industries. Most studies of SCP are based on cross-sectional data rather than data on a particular industry over time. There are two serious shortcomings in cross-sectional studies of the relationship between structure and performance across different industries.

First, it is unrealistic to expect the same relationship between structure and performance to hold across all industries. Suppose that one monopolized industry has a high elasticity of demand, and another monopolized industry has a low elasticity of demand. As Equation 8.2 shows, the price-cost margin in the industry with the high elasticity of demand is lower than the price-cost margin in the industry with the low elasticity of demand. Most cross-sectional studies fail to control for differences in demand elasticities across industries, thereby implicitly assuming that the elasticities are identical across industries.

Second, it is unlikely that the four-firm concentration ratios published by the U.S. Census Bureau correspond to the concentration ratios for relevant economic markets. If concentration ratios are not defined for the proper markets, one should not expect to find any correlation between performance and concentration across different markets.

To remedy these two problems, some studies focus on a single industry over time or across different locations. One can, for example, examine how performance in the industry changes over time because of changes in government regulation of entry. Two industry studies are reviewed here.

Airlines. The airline industry would appear to have low costs of entry between city pairs for airlines already in operation. All that is needed is to fly a plane from wherever it is to the new origin and destination pair. That is, the airline industry appears to be a contestable market. Despite the apparent ease of entry, however, studies of the airline industry consistently show that concentration in a city-pair market does influence fares.¹⁸ Actual entry, not potential entry, is critically important in influencing airline fares.

Call and Keeler (1985), Bailey, Graham, and Kaplan (1985), and Graham, Kaplan, and Sibley (1983) found that fares are higher where concentration is high. They typically concluded that fares rise by roughly 6 percent if the four-firm concentration ratio doubles from 50 to 100 percent between two cities (Bailey, Graham, Kaplan 1985,

¹⁷The industrial organization of Japan is discussed in Caves and Uekasa (1976) and Miwa (1996).

¹⁸One interpretation of this result is that it is not so easy to construct an optimal airline network that flies passengers from "spoke" cities to "hub" cities where they can interconnect other hubs or spokes. Only in very dense markets with heavy end-to-end travel between city pairs (for example, Chicago–New York) with no interconnecting passengers (and hence no need for feeder traffic) are markets likely to be contestable. Carlton and Klammer (1983) discuss the economics of such networks. The limited numbers of gates, landing slots, and take-off slots at congested airports also limit the ease of entry.

165). Again, there is a statistically significant effect of concentration on performance, but it is of modest magnitude. Borenstein (1989) presented evidence that concentration at an airport (rather than on a particular route between two cities) can also lead to modest increases in fares.¹⁹ Bamberger and Carlton (2003) find that route and airport concentration influence fares, but that this effect is much smaller when one accounts for connecting passengers. (Moreover, they find that the creation of hubs leads to output expansion, a clear benefit to consumers.) Weiher, Sickles, and Perloff (2002) show that the markup of airline fares over marginal cost depends primarily on whether one or two firms dominate a route.

Railroads. In contrast to the apparent ease of entry by airlines, it is now so costly to build a railroad that no one is likely to enter with a new large rail system. Therefore, the number of competitors can be taken as a completely exogenous variable if one focuses on commodities that are shipped only by rail and for which truck (or other) transportation is uneconomical. Studies have estimated the relationship between railroad rates as a function of distance, tons shipped, and concentration after the railroads were deregulated under the Staggers Act of 1980 and were given greater freedom to set fares.

MacDonald (1987) estimated that a railroad facing no competition can charge rates for transporting wheat that are 18 percent higher than when there is a competing railroad. When three railroads compete, rates fall by another 2 percent. These results are statistically significant, yet they indicate that rates do not go up all that much even for dramatic increases in concentration.²⁰

Measurement and Statistical Problems. In summary, there is at best weak evidence of a link between concentration and various proxies for barriers to entry and measures of market performance. Are the theories concerning the relationship between performance and structure wrong, or are these studies flawed?

Although many SCP studies are well done, others are seriously flawed. Many of the negative findings in these studies may be due to two important problems. First, these studies commonly suffer from substantial measurement problems or related statistical problems. Second, and more important, most of these studies are conceptually flawed. Most suffer from a variety of measurement errors and other statistical problems that are difficult to correct. Many of these problems were discussed above. We analyze three additional ones here.

First, concentration measures and performance measures are frequently biased due to improper aggregation across products. Because most firms sell more than one product, any estimate of profits or price-cost margins for a firm reflects averages across dif-

¹⁹See also Hurdle et al. (1989), Borenstein (1992), Brueckner, Dyer, and Spiller (1992), and Evans and Kessides (1993).

²⁰Although 20 percent is not small, it is less than one might expect as the difference between monopoly and competition among these firms. A 20 percent price overcharge is about what a monopolist would charge if it faced a demand elasticity of -6 . The demand elasticity for rail transport of grain is believed to be considerably less elastic than -6 .

ferent products. For a firm that makes products in many different industries, aggregate statistics can be misleading. For example, the Census assigns firms to industry categories based on the primary products produced and includes their total value of production under that industry category. The Census also tabulates statistics at the product level, based on data from individual plants. Because a plant is less likely than a firm to produce several products, product-level data are preferable because such data are less likely to have an aggregation bias than industry-level data.

Second, as discussed in the sections on measuring performance and structure, the performance and structural variables tend to suffer from other measurement errors. Some researchers include variables in addition to concentration to control for such measurement problems in an attempt to reduce these biases. For example, because most price-cost margins ignore capital and advertising, some economists include those two variables in their regressions of price-cost margins on concentration. The inclusion of these additional *explanatory variables* (those used to explain the measure of performance) may not eliminate the bias if they are measured with error or determined by industry profitability. For example, researchers frequently mismeasure advertising, and advertising may be more heavily used in highly profitable industries. The proper interpretation of the coefficients of variables such as advertising is that they reflect, in part, measurement error in the performance measure and not fundamental economic forces influencing “true” price-cost margins (based on marginal costs).

Third, many studies inappropriately estimate linear relations between a measure of performance and concentration. For example, if an increase in the concentration ratio has a smaller effect on performance above a certain level of concentration, the relationship between performance and concentration will flatten and resemble an S-shaped curve. This S-shaped curve can be approximated reasonably by a straight line only if the observed levels of concentration lie in the relatively straight portion of the curve. If concentration ratios vary from very low levels to very high levels, an estimate based on a presumed linear relationship may lead to incorrect results.

White (1976) and Bradburd and Over (1982) searched for critical levels of concentration below which price is less likely to increase as concentration increases, and threshold levels of concentration above which price is more likely to increase as concentration increases. They were only partially successful in finding such a level: There appears to be some evidence of an increase in price at four-firm concentration ratios above roughly 50–60 percent.²¹

Conceptual Problems. Many SCP studies have such serious conceptual problems that it is difficult to use them to test our second key question about the relationship between performance and structure. The two most common conceptual problems concern whether long-run performance measures are used and whether the structural variables are exogenous.

²¹Bradburd and Over (1982) present evidence that the effect of concentration on an industry's performance depends on levels of past concentration. As a highly concentrated industry becomes less concentrated, price remains higher than it would if the industry had never been highly concentrated.

The theories summarized in Table 8.1 predict how long-run profits vary with market structure. They say nothing about the relationship of short-run profits and market structure. Thus, an SCP study based on short-run performance measures is not a proper test of the theories.

The length of time it takes to reach the long run differs by industry. At any moment, some industries are highly profitable while others are not. Over time, some firms exit from the low-profit industries and enter the high-profit industries, which drives rates of return toward a common level. Stigler (1963), Connolly and Schwartz (1985), and Mueller (1985) find that high profits often decline slowly in highly concentrated industries. Only by analyzing both the level of profits (or other measures of performance) and the rate at which they change can the analyst distinguish between a long-run barrier to entry and the speed with which entry occurs (see Chapter 3). Most analyses do not make this distinction. This issue may be regarded as a problem in accurately measuring performance.²²

The more serious conceptual problem with many SCP studies is that the structural variables are not exogenous. Many researchers, after finding a link between high profits (or excessive rates of return, or large price-cost margins) and high concentration ratios, infer improperly that high concentration ratios are bad because they “cause” high profits. Profit and concentration, however, influence each other. An alternative interpretation of a link between profits and concentration is that the largest firms are the most efficient or innovative (Demsetz 1973, Peltzman 1977). Only when a firm is efficient or innovative is it profitable to expand in a market and make the market concentrated. In this interpretation, a successful firm attracts consumers, either through lower prices or better products. A firm’s success, as measured by both its profits and its market share, is an indicator of consumer satisfaction, not of poor industry performance. One implication of this hypothesis is that a firm’s success is explained by its own market share and not just by industry concentration, as found by Kwoka and Ravenscraft (1985).

If concentration is not an exogenous measure, then an estimate of the relationship between profits and concentration, which assumes that concentration affects profits and not vice versa, leads to what is referred to as a simultaneity bias. Weiss (1974), however, estimated the relationship between performance measures and concentration using statistical techniques designed to eliminate the simultaneity bias problem and found that the different estimation procedures make little difference in the estimated relationship.

Although the regression results may not change, their interpretation does. Even a correctly estimated relationship between performance and concentration is uninformative regarding causation. Concentration does not cause high profits; long-run barriers to entry do. These barriers lead to both high profits and high concentration.²³

²²The various measurement problems with performance may not be as serious as they first appear. Schmalensee (1989) used 12 different accounting measures of profitability in a SCP study. Strikingly, although these 12 measures are not highly correlated, many of his key SCP results held over all measures.

²³Research on SCP continues. Noteworthy work includes Marvel (1978), Lamm (1981), Cotterill (1986), Schmalensee (1987, 1989), Cubbin and Geroski (1987), and especially Sutton (1991, 1998).

Modern Structure-Conduct-Performance Analysis

The original structure-conduct-performance literature sought to establish a systematic relationship between price and concentration. As we have noted, the criticisms of this approach are many, but perhaps the most significant criticism is that concentration itself is determined by the economic conditions of the industry and hence is not an industry characteristic that can be used to explain pricing or other conduct. The barrage of criticism has caused most research in this area to cease. But Sutton and his coauthors have developed an approach that builds on the structure-conduct-performance idea of looking for systematic patterns of competitive behavior across industries, and that at the same time addresses the endogenous determination of entry (Sutton 1991, 1998).

Sutton's research examines what happens to competition as market size grows. Does the market become less concentrated? Do other dimensions of the product—such as quality, promotional activity, and research and development—change? What are the fundamental economic forces that provide the bases for systematic answers to these questions across different industries? In answering these questions, Sutton analyzes markets in which the product is either homogeneous or heterogeneous and considers the cost of entering the market or altering certain attributes of products.

Theory

We divide our discussion of Sutton's theory into two cases depending on whether a firm's cost of entry is an exogenous sunk cost or an endogenous sunk cost. In the former case, each firm must spend some fixed amount, F , to enter the industry. In the latter case, the amount a firm must spend to enter the industry is variable and is chosen by the firm in an effort to affect the desirability of its product by influencing certain dimensions of the product.

Exogenous Sunk Cost. To illustrate his theory, Sutton examines markets with homogeneous and heterogeneous products. We start by considering a market in which the firms produce a homogeneous product and the only variable firms can compete on is price, not quality. Each firm incurs a fixed cost F and has a constant marginal cost m . At low prices, the industry demand curve is $Q = s/p$, where Q is industry quantity, s is a measure of market size (total expenditure, which is assumed to be determined independently of price), and p is price. That is, for low prices and given s , the market elasticity of demand is -1 . At some high price p_m , the demand curve is perfectly elastic. Thus, a monopoly would charge a price of p_m in this market (see Chapter 4).

The final equilibrium and the change in equilibrium as the market size grows are determined by the form that competition takes. To fix ideas, Sutton considered three types of competition, each "tougher" than the next. The level of competition is lowest in a cartel in which all firms explicitly collude to set the monopoly price p_m and divide up the total cartel profit or monopoly profit among the n firms. Regardless of the

number of firms, n , the price remains at p_m . Thus, profit per firm declines as n grows because the total monopoly profit is divided among more and more firms. At the equilibrium n , the total cartel profit is driven to zero.²⁴

A more competitive market is a Cournot oligopoly. For any number of firms, n , the equilibrium Cournot price is $p(n) = m[1 + 1/(n - 1)]$.²⁵ Thus, the Cournot price p falls to m as n increases. The output per firm, q , equals $(s/m)[(n - 1)/n^2]$, while profit per firm is $[p - m]q - F$, which equals $s/n^2 - F$. Hence with free entry, n equals $\sqrt{s/F}$, at which point profit per firm is zero.

Finally, consider the toughest form of competition, Bertrand, where price equals m for any given $n > 1$. Here, the only free-entry equilibrium has one firm with positive profit. If a second firm enters, price is driven to marginal cost, so that profit is negative (because of the fixed cost), which leads to one firm's demise.

For each model of competition, Figure 8.1 shows how price changes as n increases. As the figure illustrates, for any given $n > 1$, price is lower as competition becomes "tougher," with Bertrand being the toughest and cartel being the least tough model of competition.

Figure 8.2 relates a measure of equilibrium industry concentration, $1/n$, to market size s for each model of competition, where by equilibrium market concentration, we mean that n such that total profit equals zero (or more accurately, if one additional firm enters, it will earn a negative profit).

Figure 8.2 reveals two interesting results. First, as expected, concentration falls as market size increases for all but the most competitive game (Bertrand). The intuition for this result is that larger markets can accommodate more firms.

The second result is counterintuitive: For any given market size, equilibrium market concentration is *higher*, the tougher the competition. Concentration is lowest for the cartel model, even though the cartel model has the highest price. The reason for this result is that tough competition leads to a low price, which discourages entry. This result illustrates that relying on concentration alone to make inferences about price and competitiveness can lead to erroneous conclusions.

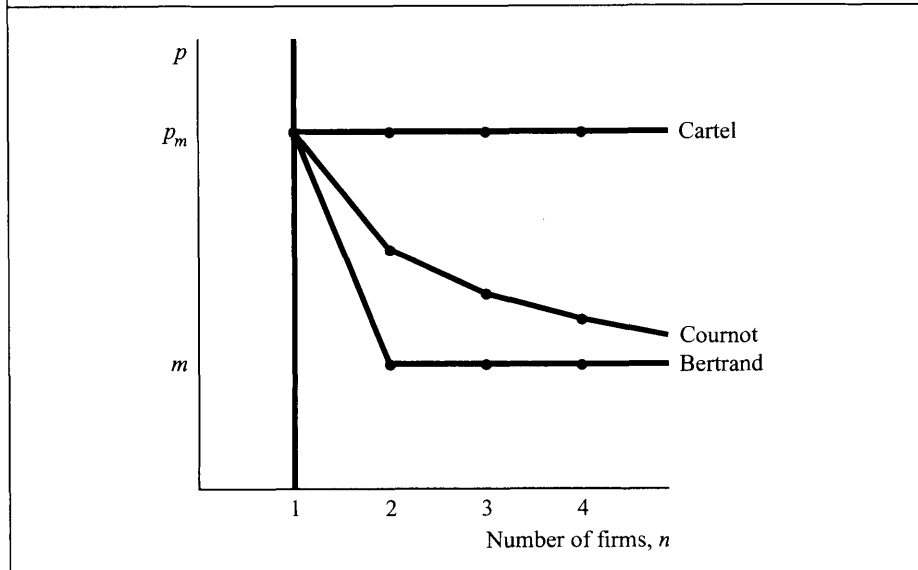
The case of exogenous fixed costs with heterogeneous products has much less crisp results than the case of exogenous fixed costs with a homogenous product. In a model with heterogeneous products (such as the models in Chapter 7), the concentration in the market depends on the nature of the game, such as how many different products one firm may produce and whether a firm has an advantage if it can choose its products before other firms choose.

²⁴Let the cartel profit be $\pi = [p - m] Q - nF$. The price that maximizes cartel profit is the same price that maximizes $[p - m] Q$. Define π_m as the maximum of $[p - m] Q$ (that is, it is the profit, ignoring fixed costs). Then, each firm's individual profit is $\pi_m/n - F$. In equilibrium where $\pi = 0$, the equilibrium n equals π_m/F .

²⁵Each firm selects its output q_i to maximize its profit, which can be written as $p_i(\Sigma q_j)q_i - mq_i - F$, where $p_i(\Sigma q_j)$ is the inverse demand curve. Differentiating this expression with respect to q_i yields the first-order condition for each firm's optimal output level given its rival's output levels. Setting $q_i = q$ for all i yields the symmetric Cournot equilibrium (assuming that the resulting price is less than p_m). See Problem 6 at the end of the chapter.

FIGURE 8.1

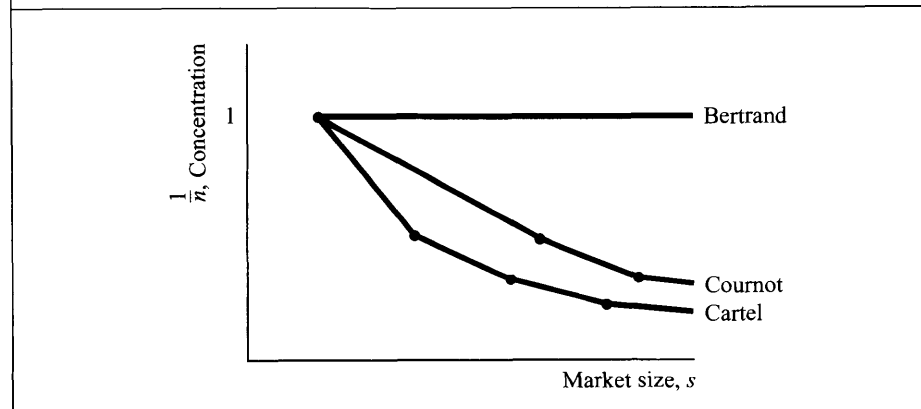
Relationship Between Prices and Number of Firms Under Three Market Structures



Sutton's main result for heterogeneous products is that the "toughness" of competition is, in general, diminished when one moves from a homogeneous to a heterogeneous product and so (analogous to the result that occurs in the case of a homogeneous product as competition weakens) the equilibrium concentration tends to fall

FIGURE 8.2

Relationship Between Concentration and Market Size Under Three Market Structures



for any given market size s . However, unlike the case of a homogeneous product, there are many possible equilibrium outcomes for any given market size, s , and the best that an economist can derive is a *lower bound* on concentration for any given s . When this lower bound is low, there are very few empirical predictions one can make about equilibrium concentration because any equilibrium concentration is possible as long as it exceeds the lower bound.

The property that equilibrium concentration (or its lower bound) decreases with market size s depends on the assumption that fixed costs are exogenous and that product quality is given. Given this property, all else equal, concentration should be lower in big countries than in small countries, when market size is determined by the size of country.

Although this result holds for many industries, there are some industries that are highly concentrated in both large and small countries (Pryor 1972). How can this fact be explained? In examining this question, Sutton and his coauthors have substantially increased our understanding of the competitive process. We now turn to their findings.

Endogenous Sunk Costs. In most markets, firms compete not just on price but also along many other product dimensions, such as quality, reliability, research and development, and promotional activity. To fix ideas, let W be an index of quality, which we will broadly interpret to include information about the product. The key new assumption is that a firm may spend money to improve its product's W . For example, firms increase their expenditures on advertising, research and development, and engineering to increase W by raising the quality of the product or by heightening consumers' perception of its quality. Firms can compete for customers by spending money to improve product quality, lowering price, or both. Here we say that the firm has *endogenous* sunk costs because the firm decides how large an investment to make.

Paying to improve quality has two important effects. First, it raises the firm's fixed cost and perhaps its marginal cost of production if a higher-quality good costs more to produce. Second, it attracts customers who were previously buying a lower-quality good. These two effects can combine to completely reverse the results in the previous section, in which an increase in market size is associated with a decrease in equilibrium concentration. As market size s increases, firms have an incentive to compete by improving the quality W of their product. To raise quality, a firm must incur larger sunk costs, a circumstance that reduces the incentive for additional firms to enter the industry that otherwise arises from the larger s . As a result, as market size increases, concentration no longer necessarily falls. A given industry in different-size markets can remain highly concentrated, but bigger markets will have higher-quality products.

For this reasoning to hold, several assumptions need to hold. Consumers must value improvements in quality sufficiently so that they switch from lower-quality products to higher-quality goods. To establish the condition under which this assumption is so, Sutton uses a model of *vertical differentiation*. In this model, every consumer agrees on a ranking of products by quality, W , with all consumers preferring a higher-quality product to a lower-quality product.

Suppose that a consumer's surplus from a good of quality W equals $U = \theta W - p(W)$, where θ is a parameter reflecting the weight that consumers place on quality

and $p(W)$ is the price of a product with quality W . Because consumers differ in θ , even though all consumers prefer more W to less, some consumers place such a low value on extra W that they are willing to pay very little extra money for a high W product, while other consumers so enjoy extra quality that they will buy a higher-quality good even if the price is relatively high. The optimal W for any consumer will depend on the price function $p(W)$, which reveals how prices rise as W rises and on the consumer preference, θ , for quality.

Sutton proves that as long as $p(W)$ and the marginal cost of producing a high-quality product do not rise “too fast” as W increases, then the equilibrium has three striking properties. First, the firms that produce the highest quality available in the market are the largest firms.

Second, an increase in market size leads to an increase in the quality of the best products in the market, with higher-quality products being chosen by consumers at higher prices and some lower-quality products disappearing from the market. Thus, the equilibrium quality rises as the market expands. Third, with higher quality and its attendant costs, fewer firms can afford to remain in the industry and concentration will remain high. Consequently, the property that a market remains concentrated as s increases continues to hold even where there is both horizontal and vertical differentiation as long as there is sufficient substitution between the vertical dimension (quality W) and the horizontal dimension over which consumers can have different preferences.

For both the endogenous and the exogenous sunk cost cases, the key empirical predictions about concentration and market size depend on the validity of certain assumptions. The most important assumption is that the form of the game—Bertrand, Cournot, or cartel—remains unchanged as market size increases. In a given market, this assumption may or may not be plausible. Moreover, neither Sutton nor anyone else has made significant progress in defining the industry economic characteristics that predict the form of the game that describes the competitive process. Therefore, analogous to the criticism of the earlier literature that concentration need not be exogenous, here we have the criticism that the form of the competitive game need not be exogenous.

Empirical Research

Sutton has produced two voluminous books of studies using data from six countries—France, Germany, Italy, Japan, the United Kingdom, and the United States—to test his theories, especially those concerning the endogeneity of advertising and technology. Sutton’s empirical work helps explain why concentration is similar across different-size countries for some industries but not others. See Example 8.1.

In Sutton (1991), he tests his theoretical predictions about the relationship between concentration and market size for several industries in the food and beverage sector. He separates the industries into two types, one in which there is little advertising and the other in which there is significant advertising. The first industry type corresponds roughly to the use of exogenous sunk cost, while the second corresponds roughly to the case of endogenous sunk cost. For each type of industry, Sutton runs a regression

EXAMPLE 8.1*Supermarkets and Concentration*

Ellickson (2000) applies Sutton's theory to the supermarket industry. In contrast to Sutton's focus on sunk costs associated with advertising or technology, Ellickson examines the role of sunk cost at the store level of building a large store and at the firm level of having the expertise and distribution systems to provide a wide variety of brands. Ellickson explains that these costs are important in distinguishing high-quality firms from low-quality stores, where he uses store size, existence of a deli or bakery, and existence of scanners and ATM machines to measure quality.

Ellickson examines the four-firm concentration ratio for supermarkets across 320 different metropolitan statistical areas (MSAs). Regardless of the size of the MSA, four or five firms that typically own multiple stores account for 70 to 80 percent of sales within each MSA. Moreover, concentration at the metropolitan level has remained high both over time and as the markets grew. Ellickson explains these results using Sutton's endogenous sunk cost theory. According to that theory, increases in market size should lead to higher-quality supermarkets but not more firms, which is exactly what Ellickson finds.

Further, consistent with the theory, the largest firm in each MSA provides a higher quality product than do the smaller firms. In addition, the quality of these largest firms differs across MSAs in exactly the way that the theory would predict: The firms in the largest markets have the highest quality.

Ellickson also examines how the industry has changed over time as MSAs have grown. The trend in the supermarket industry has been one of increasing concentration over time. For example, the average four-firm concentration ratio across 154 MSAs has grown from 45 percent in 1954 to 75 percent in 1998. Consistent with the theory, the number of products offered by each store has increased from 14,145 in 1980 to 21,949 in 1994, while average store size has been growing at the rate of 1,000 square feet per year.

of the form $C4 = a + b \ln(s/\sigma)$, where $C4$ is the four-firm concentration ratio and s/σ is the market size divided by the size of an efficient plant.²⁶ An econometric test of the theory is that b is negative for the first type of industry but zero for the second type. For his sample, Sutton indeed finds this result, which provides impressive empirical support for his theories.

Thus, Sutton's work increases our theoretical and empirical understanding of the relationship between concentration and competition. Still, there are two important caveats to Sutton's results. First, as his detailed analysis of each industry in each country reveals, the assumption that the competitive game is the same across

²⁶Sutton actually uses a more complicated method because his theory predicts a lower bound to the relationship between concentration and market size.

countries is not always a particularly good one. There is little research so far explaining why in some countries competition in a particular industry is more intense than in others.²⁷

Sutton uses the difference across countries in the competitive game for an industry to his advantage. Sutton identified industries and countries where competition is unusually intense and found, consistent with his theory, that the industry is more concentrated in those countries. He identifies countries with lax attitudes toward cartels and again, consistent with his theory, finds that those industries tend to have lower concentration levels.

The second caveat is that Sutton's theory predicts a lower bound to the relationship between market size and concentration. The reason for the lower bound is that there can be a multiplicity of equilibria, with some having greater concentration levels than the lower bound. The theory therefore is unable to help us much in predicting concentration in a particular country when the lower bound is low.

Although Sutton explains that this theory of lower bounds is the most one can say under general conditions, the analyst is left in the uncomfortable position of having a theoretical structure that may not narrow the possible equilibria very much. Sutton's detailed history of each industry shows that many idiosyncratic factors often are critical in explaining an industry's evolution. Thus his work provides a sobering lesson because it reveals the limits of theory to explain industrial structures.

● Modern Approaches to Measuring Performance

An economist's guess is liable to be as good as anybody else's.

—Will Rogers

The SCP studies focus on our second question, concerning the relationship between performance and structure, and pay relatively little attention to the first question—how to measure performance. In contrast, most modern empirical approaches focus on measuring performance or market power. These studies start by rejecting the traditional measures of performance on the grounds that they are significantly flawed due to accounting difficulties. These approaches estimate market power using models based on formal theories of profit-maximizing behavior described in earlier chapters.

Researchers use both static and multiperiod models to estimate market power. Some economists rely directly or indirectly on observations of marginal cost and price; others look at the behavior of output or price to see if it is consistent with the competitive model. The following sections discuss some of these methods.

²⁷One curious finding is that concentration tends to be slightly higher in the United States than in European countries. One explanation is that the United States has more intense competition, which Sutton's theory suggests should lead to higher levels of concentration.

Static Studies

Most modern studies based on static models can be divided into those that estimate marginal cost directly, those that estimate entire models of a market (thereby obtaining estimates of marginal cost and of the markup), and those that observe the relationship between changes in price and factor costs to test whether an industry is competitive.

Estimate Marginal Cost Using Cost Data. The most direct way to answer our first key question about the degree of market power in an industry is to calculate the price-cost markup directly.²⁸ Although price data are available for most industries, unfortunately, marginal cost data are generally not.

If information on total cost is available, however, an economist can estimate the relationship between observed total cost and observed total output and then calculate marginal cost. A price-cost margin is then simply calculated. Weiher et al. (2002) estimated marginal cost using total cost information and then calculated Lerner measures of market power directly.

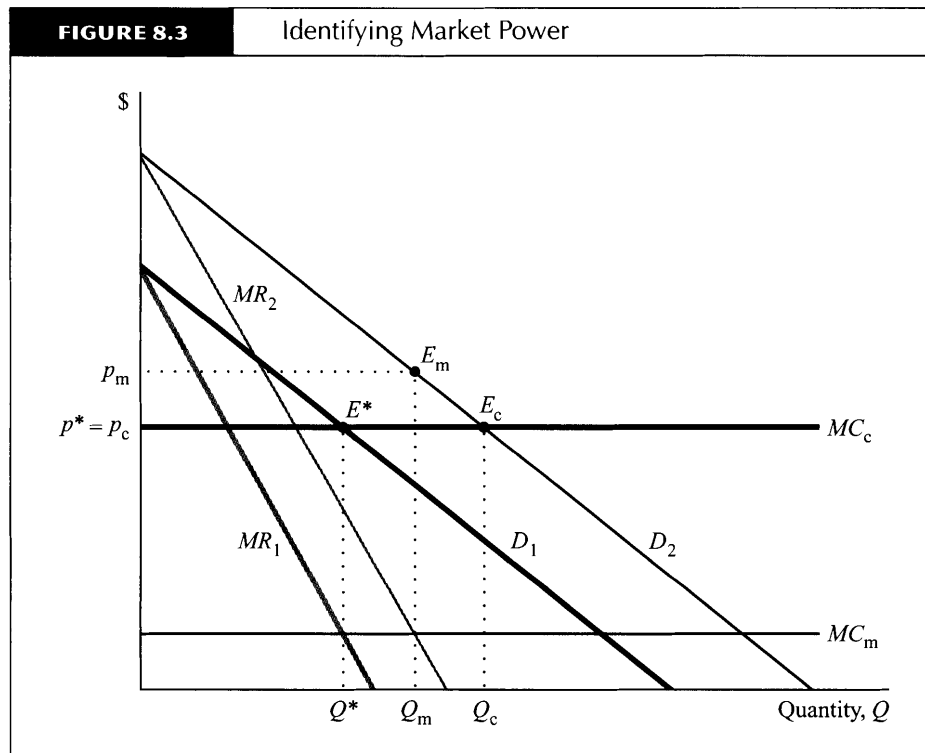
Even total cost data, however, are rarely available. Studies that estimate cost functions frequently examine regulated industries because regulators force the firms to provide cost data. For example, Keeler (1983, 71) and Friedlaender and Spady (1980, Ch. 4) found that price exceeded long-run marginal cost by about 22 percent for rail service for bulk commodities in the Northeast during the late 1960s and early 1970s. Genesave and Mullin (1998) use cost data from a court case and find small markups.

Estimate the Markups Using an Industry Model. If cost data are not available, so that we cannot directly estimate marginal cost, MC , how can we calculate the price-cost markup? One method is to use assumptions about the shape of the demand and MC curves to infer the markup from observations on how the equilibrium price and quantity change over time.²⁹ This approach is called the new empirical industrial organization.

For many markets, we have enough information to estimate a demand curve. Figure 8.3 shows the demand curve, D_1 , in a particular market. Suppose that we believe that the industry's marginal cost, MC , is constant, although we do not know its level. Currently, the market equilibrium, point E^* in Figure 8.3, is at price p^* and quantity Q^* . That equilibrium could be produced by a competitive industry with a relatively high marginal cost, MC_c , or by a monopoly with a relatively low marginal cost, MC_m .

²⁸Hall and Hitch (1939) conducted a series of interviews with businesspeople regarding their firms' pricing practices. Most claimed that they set price above marginal cost.

²⁹See Bresnahan (1989) for a more extensive discussion and Corts (1999) for a critique that explains that these empirical methods depend on the validity of the conjectural variations model that is theoretically an inadequate model. Apparently the first modern study was Rosse (1970). Six other influential early studies are Iwata (1974), Applebaum (1979, 1982), Gollop and Roberts (1979), Just and Chern (1980), and Bresnahan (1981). The other major early conceptual work was Rohlfs (1974); however, it did not contain an empirical application.



(which intersects the monopoly's marginal revenue curve, MR_1 , at Q^*). With only this information, we cannot *identify* (determine) the marginal cost and price-cost markup.

However, if in the next period the demand curve shifts to D_2 , which is to the right and parallel to D_1 in Figure 8.3, we can determine whether the industry is competitive or monopolistic. If the industry is competitive, the new equilibrium is at point E_c , so the price remains constant, $p_c = p^*$, and output increases substantially to Q_c . That is, by noting that the shift in demand does not change the price, we know that the industry marginal cost is MC_c and that Lerner's price-cost margin, $(p - MC_c)/p$, is 0.

If, instead, the shift in demand leads to a new equilibrium at point E_m , the price increases from p^* to p_m and the quantity increases only to Q_m . This increase in price is consistent with noncompetitive behavior. Thus, if we know MC is constant, an outward shift of the demand curve reveals whether the market is purely competitive. If the price does not change, the market is competitive; if the price increases, there is market power.

Economists use generalizations of this approach to estimate the degree of market power, Lerner's price-cost margin, and the marginal cost curve. Typically, they make specific assumptions about the shapes of the demand and marginal cost curves, which allow them to identify the price-cost margin by observing shifts in equilibrium price and quan-

TABLE 8.7 Estimated Price-Cost Margins

Study	Industry	$(p - MC)/p$
Bresnahan (1981)	Autos	.10-.34
Appelbaum (1982)	Rubber	.05
	Electrical machinery	.20
	Tobacco	.65
Porter (1983a)	Railroads (with cartel)	.40
Lopez (1984)	Food processing	.50
Roberts (1984)	Coffee roasting (largest firm)	.06
Spiller and Favaro (1984)	Large banks before deregulation	.88
	Large banks after deregulation	.40
Suslow (1986)	Aluminum	.59
Slade (1987b)	Retail gasoline	.10
Karp and Perloff (1989a)	Rice exports (largest estimate)	.11
Karp and Perloff (1989b)	Small black-and-white TVs in Japan	.58
Buschena and Perloff (1991)	Philippines coconut oil (post-1974)	.89
Wann and Sexton (1992)	Fruit cocktail	1.41
Deodhar and Sheldon (1995)	German bananas	.26
Genesove and Mullin (1998)	Sugar refining 1880-1914	.05
Hyde and Perloff (1998)	Australian retail meats	≈ 0

Sources: Articles cited and Bresnahan (1989, Table 1)

tity over time. One method is described in Appendix 8B.³⁰ Table 8.7 reports the estimated price-cost margins for several industries using these approaches. See Example 8.2.

Indirect Approaches. Some economists use the changes in price associated with changes in costs to test whether an industry is competitive without having to make detailed assumptions about the shapes of both the demand and the supply curves. If marginal cost shifts up by a certain amount in a constant-marginal cost market, the competitive price rises by the same amount because price equals marginal cost. For ex-

³⁰Bresnahan (1989) surveys many of these studies, including Iwata (1974), Gollop and Roberts (1979), Spiller and Favaro (1984), Roberts (1984), and Applebaum (1979, 1982). Analogous techniques can be used to estimate monopsony power as well (Just and Chern 1980, Azzam and Pagoulatos 1990). Separate price-cost margins can also be estimated for individual firms, as shown in Spiller and Favaro (1984), Baker and Bresnahan (1985, 1988), Slade (1986, 1987a, 1987b, 1992), Gelfand and Spiller (1987), and Karp and Perloff (1989b). Any shock that shifts the relevant demand curve or marginal cost curve can be used to identify market power. For example, changes in taxes (Kolstad and Wolak 1983, 1985, 1986; Wolak and Kolstad 1988) or changes in the supply of a fringe (Buschena and Perloff 1991) help identify market power. Some of the more interesting applications (Bresnahan 1981, 1987) estimate market power based on spatial competition models (Chapter 7), taking explicit account of product differentiation.

EXAMPLE 8.2*How Sweet It Is*

If we have data on marginal cost, MC , and market price, p , we can calculate the Lerner Index (Chapter 4) of market power, $(p - MC)/p$, directly. Unfortunately, such cost data are usually not available.

As a result, most new empirical industrial organization studies of market power identify the Lerner Index by using assumptions about the shape of the demand and marginal cost curves. Consequently, these estimates of market power are only as reliable as the (untested) assumptions about the shapes of the curves. Typically, these studies estimate the degree of market power using a parameter λ so that the Lerner Index is (Appendix 8B)

$$(p - MC)/p = -\lambda/\epsilon,$$

where ϵ is the estimated market demand elasticity. If the market is competitive, $\lambda = 0$ and there is no gap between price and marginal cost. If the market is monopolized, $\lambda = 1$. If λ is between zero and one, the degree of market power is between that of a competitive and monopolized market.

Genesove and Mullin (1998) have data on cost for the sugar refining industry. Consequently, they are able to see how well the estimation approach works compared to directly calculating market power using cost data.

The sugar refining industry became a highly concentrated industry as the result of acquisitions by the American Sugar Refining Company in the late 1800s. The largest firm accounted for over 60 percent of sales. Detailed cost data for this industry for 1880–1914 are available as a result of antitrust litigation. Genesove and Mullin use these data to calculate marginal cost directly, which they then use to calculate the Lerner Index. According to these calculations, a typical value of the Lerner Index during the 1880–1914 period is .05 (nearly perfectly competitive), while a typical value for λ is .1.

Next, Genesove and Mullin ignore their detailed cost data and use a technique similar to that described in Appendix 8B to estimate the demand curve, the marginal cost curve, and λ . Using this method, they estimate that $\lambda = .04$ and that the implied Lerner Index is .02. Although their econometric approach leads to a lower estimate of λ than does their cost method, the econometric method succeeds in correctly telling the researchers that, despite the high industry concentration, a monopoly ($\lambda = 1$) model is less consistent with the data than is a competitive ($\lambda = 0$) model.

ample, in a competitive market, a per-unit tax of \$1 raises price by \$1. By observing the relationship between the change in price and the change in costs (or some element of costs), one can test whether the industry is competitive.

Sumner (1981) examined the effect of tax differences across states on the price of cigarettes. He argued that if the retail prices of cigarettes differ between states by the amount of the tax differences, the market is relatively competitive. Bulow and Pfleiderer (1983) pointed out that it is possible to construct demand curves for which a monopoly does pass on costs on a one-for-one basis. Sullivan (1985) used a different method to avoid this criticism and confirms Sumner's finding of a significant degree of competition in cigarettes. Similarly, Ashenfelter and Sullivan (1987) used changes in excise taxes to identify market structure.

Hall (1988a) demonstrated another method of determining market power without making specific assumptions about the demand curve. He showed that, with constant returns to scale, shifts in costs are sufficient to identify market power.³¹ When such an industry expands output in response to a shift in demand, the total value of its output (revenues) increases by exactly the increase in its total cost if the industry is competitive. If value rises by more than the additional cost, then price is above marginal cost and the industry is not competitive.³²

Hall estimates very large markups, but subsequent work by Domowitz et al. (1988) and Roeger (1995) find much lower markups.³³ Roeger (1995) obtains markups ranging from 5 to 23 percent.

Multiperiod Studies

Almost all real-world markets last for many periods. A multiperiod model should be used to estimate market power if firms, in setting strategies, take previous behavior into account; if adjustment costs are significant, so that costs in this period depend on decisions in previous periods; or if demand today depends on past consumption. Economists use at least two types of multiperiod models to estimate market power: models of collusive behavior and models of behavior with costs of adjustment.

Collusion and Repeated Static Games. Stigler (1964a) argued that the opportunity and desire by oligopolistic firms to collude (at least tacitly) provides a basis for explaining all oligopoly behavior (Chapter 5). In this theory, prices below the monopoly

³¹Rosse and Panzar (1977), Panzar and Rosse (1987), and Shaffer (1982) showed how to test whether a market is competitive, oligopolistic, or monopolistic using information on shifts in revenue in response to shifts in factor prices. To estimate the actual degree of market power, however, one must have additional information or make some strong assumptions such as Hall's constant returns to scale assumption.

³²Suppose that the industry has a demand curve with a constant elasticity of ϵ and a constant marginal cost. A monopoly sets price equal to $1/(1 + 1/\epsilon)$ times the constant marginal cost (as can be shown by rearranging Equation 8.2). If ϵ is -2 , then the price is twice the marginal cost. If, holding ϵ constant, demand shifts out so that one more unit is sold, revenues rise by p , but total cost increases by MC , which is only half of p .

³³Domowitz et al. (1988) do not find that concentration plays an important and statistically significant role in explaining the deviation between price and marginal cost. However, Shapiro (1987), using a variant of Hall's method, does find a strong relation between margins and concentration.

level are due to failures to enforce the cartel fully. In this story, market structure matters. For example, the more firms in an industry, the harder it is to detect cheating by any one firm, so more cheating occurs, and the average price is lower.

Game theorists model Stigler's insight as a supergame over repeated static games. In one version, random fluctuations in price due to fluctuations in demand or supply costs could make "cheating" by cartel members hard to detect because the price fluctuations could be due to either cheating or shifts in economic conditions. To prevent firms from cheating, all cartel members agree that if the market price drops below a certain level—a "trigger price"—each firm will expand its output to the precartel level for a certain period of time and prices will fall as a result. If firms expect other firms to stick to this agreement, a firm that cut its price might gain in the extremely short run, but would lose in the end because of the destruction of the cartel by this predetermined punishment mechanism (Chapter 5).

Porter (1983a), Lee and Porter (1984), and Ellison (1994) used this theory to estimate a model of 1880s railroad cartel behavior. Comparing high and low price periods, Porter finds that the cartel increased its rate by over 60 percent during periods of successful collusion. See Example 5.6.

Dynamic Models with Adjustment Costs. If firms have substantial adjustment costs from training new workers, from storage of inputs or outputs (inventories), or in accumulating capital, they must plan their actions over many periods if they are to maximize long-run profits. For example, if the firm must pay compensation to laid-off workers (an adjustment cost), the firm hires fewer workers in period t if it believes demand will be lower in period $t + 1$. Similarly, firms' costs may fall over time if there is **learning by doing** (costs fall with production because workers become more skilled at their jobs due to experience or as better ways of producing are discovered); actions by a firm in this period affect its costs and profits in later periods.³⁴

Pindyck (1985) showed that, in a dynamic setting, a mechanical application of the Lerner Index for each period can be misleading. In the intertemporal case, neither the short-run demand elasticity nor the Lerner Index provides a meaningful measure of monopoly power. One solution is to discuss the steady-state price-cost margin (the margin that eventually would be reached and that would persist if there were no further cost or demand shocks) or to compare the path of price or quantity with respect to the path under the price-taking assumption.

The game-theoretic literature abounds with dynamic models of oligopoly that are too general to be usable in estimation. To estimate these models practically, further restrictions have to be imposed. Roberts and Samuelson (1988) use a dynamic oligopoly model with reasonably general functional forms to reject the hypothesis that the ciga-

³⁴Analogous to dynamic models with adjustment costs are those where demand today depends on quantities in previous periods. Some marketing studies attempt to estimate demand curves with this property, as do studies of durable goods such as aluminum (Suslow 1986b). Similarly, in pumping oil, the costs today depend on how much was pumped in the past and price is expected to rise at the rate of interest (according to the Hotelling formula), so empirical studies of oil reflect these dynamic issues as well (Matutes 1985).

rette market is competitive. With their general functional form, however, they cannot estimate the degree of market power. Karp and Perloff (1989a, 1993a) used a dynamic oligopoly model with a linear demand curve and quadratic costs of adjustment to estimate steady-state price-cost margins for the international coffee and the international rice export markets. For recent work on dynamic oligopoly, see the references cited in Chapter 6, especially Ericson and Pakes (1995, 1998), Fershtman and Pakes (2000), and McGuire and Pakes (1994).

Value of Modern Approaches to Measuring Performance

The modern approaches have three major advantages over the SCP approach. First, they estimate the market performance rather than use an accounting proxy. Second, they use changes in exogenous variables (wages, taxes, demand growth) to explain variations in performance rather than endogenous variables such as concentration ratios and advertising. Third, they are based on maximizing models for individual industries so that hypotheses about behavior can be tested. Their key disadvantage is that many of these models require making detailed assumptions about the shapes of the supply and demand curves and about oligopoly behavior. Moreover, none of the modern approaches that we have discussed focuses on the use of cross-sectional variation across industries to make any predictions as to what factors cause competition to differ across industries. It was the search for such factors that was at the heart of the SCP approach and central to Sutton's approach.

SUMMARY

The empirical relationship between measures of performance, such as price-cost margins, and market structure, such as concentration and entry barriers, is not clear. Serious measurement problems can plague such structure-conduct-performance (SCP) studies. Accounting measures of performance may fail to measure economic profits or costs accurately, especially when long-lived capital assets are present. Concentration ratios for individual industries can be measured accurately, but make sense only when the individual industries constitute a relevant economic market. Finally, the measurement of barriers to entry is often subjective and typically fails to distinguish between long-run barriers to entry and the speed with which entry can occur.

Studies relating measures of industry performance to concentration and barriers to entry across industries suffer from several conceptual problems. A statistically significant relationship between concentration and performance would not necessarily imply that concentration caused price to be above the competitive level. An alternative explanation is that firms become large (concentration rises) because they are efficient. If so, within an industry, profits of the largest firms are higher than those of the smallest. The empirical results indicate either no effect or a small positive effect of concentration and barriers to entry on performance, but this effect is often statistically insignificant. Sutton and his collaborators have produced research that addresses many of the criticisms of the SCP approach and simultaneously uses industry information to make predictions about industry concentration.

Studies of individual industries can avoid many, though not necessarily all, of the conceptual problems of older SCP cross-sectional studies. Such studies tend to find a small but statistically significant effect of concentration on industry measures of performance, such as price.

Modern studies statistically estimate the price-cost margin for a particular industry rather than rely on accounting proxies. These studies have their own disadvantages: Researchers typically have to make detailed assumptions about demand, cost functions, or oligopoly behavior. Many of these industry studies find substantial margins. These methods have not yet been used to explore in detail the relationship of industry structure to the degree of deviation from perfectly competitive behavior.

PROBLEMS

1. Why do empirical researchers often include the advertising-sales and the capital-sales ratios in equations explaining performance?
2. An industry has a price of p^* and earns a rate of return r^* on its capital. The industry is characterized by a fixed-proportions production technology (a fixed proportion of labor and capital is required to make each unit of output—no substitution between labor and capital is possible). Let p and r be the price and rate of return, respectively, that would emerge if the industry were competitive. What is the relationship between $p^* - p$, $r^* - r$, and the capital-output ratio?
3. Concentration ratios are typically a firm's share of *domestic* production. If the United States engages in more international trade, will such concentration measures lose meaning? Could this effect explain the vanishing of the price-concentration effect over time?
4. (*Difficult*) Evaluate the following argument: "There exist demand curves for which a monopoly would pass along cost increases in price on a one-for-one basis. Therefore, nothing can be inferred about the competitiveness of an industry by comparing price changes to cost changes." In your evaluation, see if you can derive a demand curve with the stated properties (Bulow and Pfleiderer 1983).
5. Distinguish between zero profits and a price-cost margin that equals zero.
6. Suppose that the demand function is $Q = s/p$, where Q is the total quantity demanded, s is a measure of the size of the market, and p is the price of the homogeneous good. Let F be a firm's fixed cost and m be its constant marginal cost. If n firms compete in a Cournot model, calculate the price, p , a typical firm's output, q , and a typical firm's profit, π .
 - a. Prove that:
 - i. $p = m \left[1 + \frac{1}{n-1} \right]$,
 - ii. $q = \frac{s}{m} \frac{n-1}{n^2}$, and
 - iii. $\pi = s/n^2 - F$.
 - b. If entry is free, what does n equal?
 - c. What happens to equilibrium concentration, $1/n$, as s increases?
 - d. What happens to equilibrium firm size as s increases?

Answers to odd-numbered problems are given at the back of the book.

APPENDIX 8A

Relationship Between the Herfindahl-Hirschman Index (HHI) and the Price-Cost Margin

An oligopoly consists of n identical firms that produce a homogeneous product. Each Firm i chooses its output, q_i to maximize its profits,

$$\pi_i = p(Q)q_i - mq_i,$$

where m is the constant marginal (and average variable) cost for each firm, and p , the price, is a function of total industry output, $Q = nq_i$.

The firms play Cournot (see Chapter 6), so each firm's first-order condition—which is obtained by setting the derivative of profits with respect to q_i equal to zero—is that marginal revenue equals marginal cost:

$$MR = p + q_i p' = m = MC, \quad (8A.1)$$

where p' is the derivative of price with respect to Q . Rearranging the terms in Equation (8A.1), this expression can be expressed in terms of the Lerner Index:

$$L \equiv \frac{p - m}{p} = -\frac{p' Q}{p} \frac{q_i}{Q} = -\frac{s_i}{\epsilon} = -\frac{1}{n\epsilon}, \quad (8A.2)$$

where $s_i \equiv q_i/Q = 1/n$ is the output share of Firm i and $1/\epsilon = (p' Q)/p$ is the reciprocal of the elasticity of demand. Because all firms are identical, Equation (8A.2) holds for every firm in the industry.

As Cowling and Waterson (1976) show, the industry average of firms' price-cost margins using share weights is

$$\sum_i s_i \frac{p - m}{p} = -\frac{\sum_i s_i^2}{\epsilon} \equiv -\frac{HHI}{\epsilon},$$

where HHI is the Herfindahl-Hirschman Index. That is, the HHI divided by the absolute value of the market demand elasticity equals the weighted average of the firms' price-cost margins.

APPENDIX 8B*Identifying Market Power*

Under what conditions can the price-cost margin be determined if we cannot observe marginal cost directly? One approach to answering this question involves estimating a complete model of the market where the shapes of the demand and marginal cost curves are specified and profit-maximizing behavior is assumed.¹

To illustrate this approach, suppose that an industry consists of a number of identical firms that produce a homogeneous product. The demand curve is $p(Q; Z)$, where p is the single price in the market, Q is output, and Z is another variable that affects demand, such as income or the price of a substitute.

Because industry revenues are $R \equiv p(Q; Z)Q$, we define the effective (or perceived) marginal revenue as

$$MR(\lambda) = p + \lambda p_Q Q,$$

where λ is a parameter to be estimated and $p_Q \equiv \partial p / \partial Q$. If the industry is monopolized, $\lambda = 1$ and effective $MR(1)$ is the usual MR measure: $p + p_Q Q$. If the firms in the industry are price takers, $\lambda = 0$ and effective $MR(0)$ equals price. Various other oligopolistic and monopolistically competitive market structures produce a λ that lies strictly between 0 and 1.

The profit-maximization or optimality condition is that effective marginal revenue equals marginal cost: $MR(\lambda) = MC$. As a result, λ is a measure of the gap between price and marginal cost. That is, the Lerner's Index is

$$L \equiv \frac{p - MC}{p} = -\frac{\lambda p_Q Q}{p} = -\frac{\lambda}{\epsilon},$$

where ϵ is the market elasticity of demand. This expression is very similar to those derived in Appendix 8A that depend on the number of firms, the market share, or the Herfindahl-Hirschman Index.

As an example, suppose that the demand curve has the particular linear form

$$p = \alpha_0 + \alpha_1 Q + \alpha_2 Z + \alpha_3 ZQ + \epsilon_1, \quad (8B.1)$$

¹The following discussion of the role of market demand shocks in identifying market power is based on Just and Chern (1980), Bresnahan (1982), and Lau (1982).

so that the effective marginal revenue is

$$MR(\lambda) = p + \lambda p_Q Q = p + \lambda(\alpha_1 + \alpha_3 Z) Q. \quad (8B.2)$$

A profit-maximizing firm sets its effective marginal revenue equal to its marginal cost. If its marginal cost curve is linear in Q and factor price W ,

$$MC = \beta_0 + \beta_1 Q + \beta_2 W + \epsilon_2,$$

its optimality equation, $MR(\lambda) = MC$, can be written as

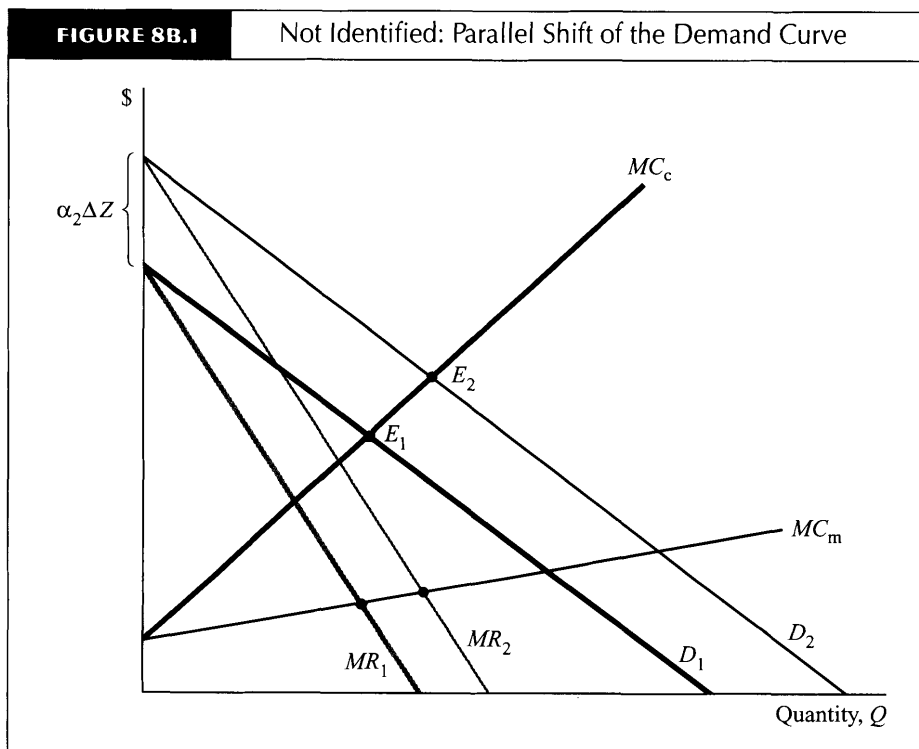
$$p = \beta_0 + (\beta_1 - \lambda\alpha_1) Q - \lambda\alpha_3 ZQ + \beta_2 W + \epsilon_2. \quad (8B.3)$$

Using the appropriate statistical techniques, one can regress p on a constant, Q , ZQ , and W to obtain estimates of the coefficients in Equation 8B.3. By dividing the estimate of the coefficient on the ZQ term, $-\lambda\alpha_3$, from Equation 8B.3 by the estimate of α_3 from the demand Equation 8B.1, one obtains an estimate of the market structure parameter λ . The reason that one can identify λ is that the demand and MR curves rotate with Z due to the ZQ interaction term, which affects where the MR curve intersects the MC curve. Alternatively, if we know MC , we can use the information about price from the demand curve to determine λ . Rotating the demand curve leaves the level of demand unchanged at the rotation point, but changes the elasticity of demand. As the elasticity of demand changes, the price changes, which allows us to estimate λ .

If there is no ZQ term (that is, if $\alpha_3 = 0$) in the demand curve, λ may not be identified. The only remaining term with a λ in Equation 8B.3 is $(\beta_1 - \lambda\alpha_1) Q$. Although we know α_1 from the demand equation, that is not enough to identify λ because the estimated coefficient also depends on β_1 (the unknown slope of the MC curve).

The need for the demand curve to rotate is illustrated in Figure 8B.1.² Initially, the researcher observes the market equilibrium, E_1 , price and quantity. The researcher estimates the demand curve D_1 (and, hence, can infer the marginal revenue curve, MR_1) but does not directly observe costs. The observed equilibrium, E_1 , is consistent with a competitive industry structure and a marginal cost curve MC_c , where the equilibrium, E_1 , is determined by the intersection of MC_c and D_1 . It is also consistent with a cartelized market structure and a lower marginal cost curve, MC_m , where the quantity associated with E_1 is determined by the intersection of MC_m and MR_1 .

²Lau (1982) shows that virtually any functional form for the demand curve leads to identification except the two most commonly used forms: linear or log-linear. If one wants to use a basically linear specification, one must add an interaction term, a squared term in output, or something else that adds some nonlinearity and allows the demand curve to rotate. Even if one does that, there is an additional serious problem with the linear specification: see Perloff and Shen (2001).

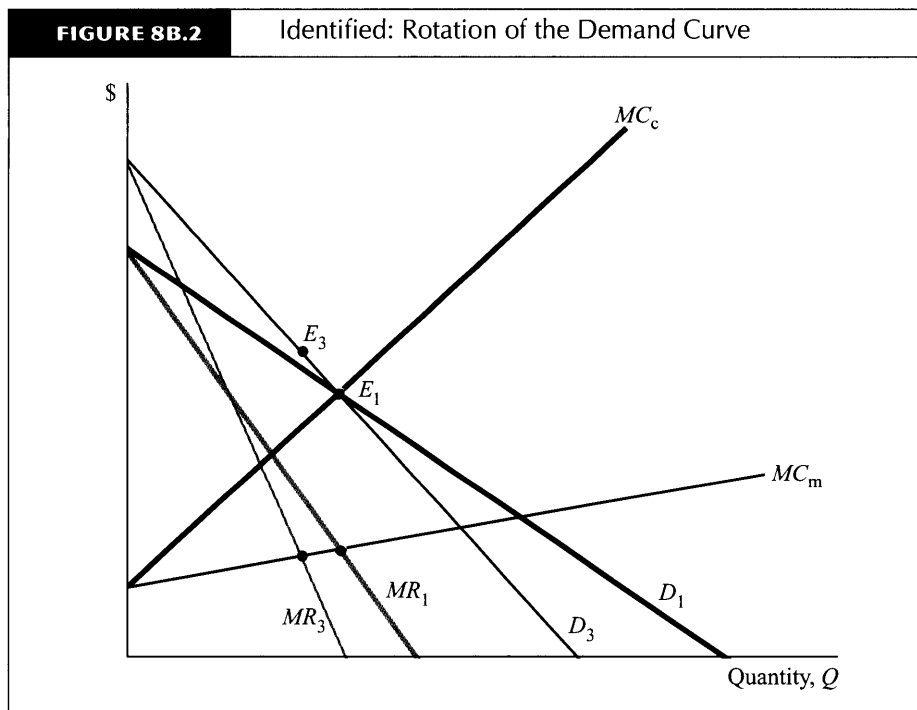


If $\alpha_3 = 0$, and Z increases by ΔZ , the intercept of the demand curve shifts up by $\alpha_2 \Delta Z$, as shown for the new demand curve, D_2 . The new equilibrium, E_2 , is still consistent with either of the two marginal cost curves. Thus, the researcher cannot determine from this shift in Z if the industry is competitive or cartelized.

In contrast, if $\alpha_3 \neq 0$, a shift in Z reveals λ . In Figure 8B.2, when Z increases, the new demand curve, D_3 , rotates (for graphical simplicity, D_3 is rotated around the original equilibrium point). If the industry is competitive and the marginal cost curve is MC_c , the new equilibrium on D_3 remains E_1 ; whereas, if the industry is cartelized and the marginal cost curve is MC_m , the new equilibrium on D_3 is E_3 . Thus, whether or not the equilibrium shifts reveals whether the market is competitive.

Anything (not just variables in the market demand curve) that causes the residual demand curve facing a firm to rotate can identify λ . For example, a dominant firm's residual demand curve is the market demand curve minus the supply of a competitive fringe. If the fringe supply curve rotates, the residual demand curve rotates even if the market demand curve does not. Similarly, a shift in an ad valorem tax rate, t , can identify the market structure.

As the chapter shows, information about the shape of the marginal cost curve also can help identify λ . It is possible to identify λ even if the demand curve does not rotate



($\alpha_3 = 0$), if the marginal cost curve is constant in Q ($\beta_1 = 0$). Because $MC = \beta_0 + \beta_2 W$, marginal cost is a constant in any given period, but that constant shifts with the exogenous factor W over time causing price to change, which allows one to estimate the demand curve. The coefficient on the Q term in Equation 8B.3 is $\beta_1 - \lambda\alpha_1 = -\lambda\alpha_1$, so by knowing α_1 from the demand curve, one can identify λ .

Antitrust Laws and Policy

The first thing we do, let's kill all the lawyers. —William Shakespeare

The U.S. government uses **antitrust laws** to limit the market power exercised by firms and to control how firms compete with each other. The antitrust laws do not make monopoly illegal, but they control how firms attain and maintain their market power. This chapter describes the antitrust laws and how they affect efficiency. It is not intended as a complete course in antitrust law; it is designed to provide an overview of the most important developments and issues in federal antitrust policy.¹

The chapter first describes the major antitrust statutes and their major objectives. As with most laws, a literal reading of the statutes does not convey how the laws have been applied. Our discussion of court decisions concentrates on market power, which is a central focus of the antitrust laws.

The chapter then examines the two major areas of antitrust law applications. The first deals with agreements among competitors, such as price-fixing agreements and agreements to merge. The second deals with the actions of a single firm that may harm rivals. These actions involve strategic behavior such as predatory pricing, vertical relationships among firms, and tie-in sales. Next, the chapter reviews the antitrust doctrines on price discrimination. The chapter concludes with an overall economic assessment of the effect of the major antitrust doctrines on firm organization.

¹See Posner and Easterbrook (1980), Posner (1976a), Bork (1978), Areeda and Turner (1978, 1980), and Areeda (1986) for more detailed examinations of antitrust issues. The presentation and analysis of the cases in this chapter rely heavily on Posner and Easterbrook (1980) and its supplements. Although states have antitrust laws, we concentrate on federal laws.

The seven main points we make in this chapter are

1. The interpretation of U.S. antitrust laws varies over time.
2. Antitrust laws should promote efficiency.
3. Monopoly is not prohibited, but certain activities that could lead to a firm's acquiring or exercising monopoly power are banned.
4. Price fixing is generally prohibited.
5. Certain agreements among competitors, vertical relations between firms, and various other strategic acts may increase or decrease welfare; hence, they should be evaluated on a case-by-case basis.
6. A few antitrust laws, such as the prohibition on price discrimination that allegedly reduces competition among customers, almost always lower welfare.
7. Prohibiting certain activities and not others can lead to inefficient organizations of firms.

The Antitrust Laws and Their Purposes

The antitrust laws are simple to state but have proved difficult to apply. Indeed, the Supreme Court has changed its interpretation of these laws several times. This section describes the laws, their enforcement, and their purposes. It explains who can sue under the antitrust laws and how damages are paid.

Antitrust Statutes

The three major statutes governing antitrust policy are the Sherman Act (passed in 1890), the Clayton Act (1914), and the Federal Trade Commission Act (1914). Additions, deletions, and amendments to these statutes have been made over the years.

Even prior to the passage of the Sherman Act, however, legal principles governed competition among firms. Under the common law (the precedents based on court decisions in the absence of explicit statutes), price-fixing among firms, though not illegal, was unenforceable: A court would not enforce a contract in which one firm agreed with a competitor to fix prices. Similarly, agreements not to compete that accompanied the sale of a business or an employment relationship were also unenforceable if they were judged “unreasonable.” Agreements among workers either to fix wages or to strike were often held to violate the law. Practices by which firms attempted to exclude competitors (for example, predatory pricing) were not considered to violate the law unless accompanied by additional illegal actions such as fraud (Posner and Easterbrook 1980, 18).

The antitrust statutes were passed at a time of great upheaval in American industry. Around 1890, when the Sherman Act was passed, large firms became increasingly common with the birth of the modern American corporation and the creation of very large firms through mergers and scale economies. Adjusting for the size of the economy, the merger wave in the 1890s and early 1900s was the largest in our history (Chapter 2).

The *Sherman Act*, the first federal antitrust legislation, was, in part, a response to these changes in the U.S. economy. Section 1 of the Sherman Act states that

every contract, combination in the form of trust or otherwise, or conspiracy, in restraint of trade or commerce among the several States, or with foreign nations, is declared to be illegal. . . .

That is, Section 1 forbids explicit cartels.

Section 2 states that

every person who shall monopolize or attempt to monopolize, or combine or conspire with any other person or persons, to monopolize any part of the trade or commerce among the several States, or with foreign nations, shall be deemed guilty of a felony. . . .

Although one might read Section 2 as prohibiting monopoly, courts have interpreted it differently. As explained later, it is not a crime to be a monopoly as long as the monopoly does not commit “bad acts.”

The courts’ interpretation of the Sherman Act left doubt as to whether the Act prohibited certain industry behavior. As a result, in 1914, legislators passed additional antitrust legislation: the Clayton Act and the Federal Trade Commission Act. The *Clayton Act* is directed primarily against four specific practices. Section 2 of the Clayton Act (amended in 1936 by the Robinson-Patman Act) prevents price discrimination that lessens competition. Section 3 prohibits the use of tie-ins and exclusive dealing when the result is to lessen competition. Section 7 (amended in 1950 by the Celler-Kefauver Act) prohibits mergers that reduce competition. Section 8 deals with the creation of interlocking directorates among competing firms (that is, the control of competing firms by interrelated boards of directors). The Clayton Act also allows an injured party to recover **treble damages** (three times actual damages) plus attorneys’ fees.²

The *Federal Trade Commission Act* created a new government agency, the Federal Trade Commission (FTC), which enforces antitrust laws and adjudicates disputes under the antitrust laws under the Federal Trade Commission Act in addition to other activities. The main antitrust provision of the FTC Act is Section 5, which prohibits “unfair” methods of competition. The FTC’s other main responsibilities include consumer protection and the prevention of deceptive advertising.

It is common for an antitrust complaint to list violations of several of the antitrust statutes simultaneously. So, for example, an antitrust complaint regarding tie-in sales could list violations of both the Sherman Act and the Clayton Act.

Enforcement

Both the FTC and the U.S. Department of Justice (Justice Department) are responsible for administering the antitrust laws. A suit brought by the Justice Department is adjudicated in federal court, whereas an action brought by the FTC is heard and decided by an administrative law judge at the FTC and then reviewed by the Federal

²A profit-maximizing firm has an incentive to violate antitrust laws if the expected punishment is less than the expected gain. If the probability of being caught is less than one (certainty), a fine equal to the damage caused may not discourage this activity. Thus, larger (treble) damages are used. See, however, Salant (1987), who argues that trebling damages may have undesirable effects if buyers anticipate receiving damage awards.

Trade Commissioners.³ After the FTC has completed its proceedings, defendants can appeal adverse decisions to the federal courts.

An action brought by the FTC can result in a *cease and desist* order, which prohibits specific acts. A suit brought by the Department of Justice can result in a similar type of order, an *injunction*. The Department of Justice can also bring a criminal suit, which may result in criminal fines or jail sentences. Aside from its enforcement responsibilities, the Department of Justice can sue to recover the cost of the suit plus the damages that arise when the U.S. government is a victim of an antitrust offense. A private individual or firm can bring an antitrust suit and, if victorious, receive treble damages plus the cost of the suit including attorneys' fees. Such private litigation comprises a significant share of antitrust litigation (White 1989).

Goals of the Antitrust Laws

Most economists believe the antitrust laws *should* have the very simple goal of promoting efficiency. That is, they should prevent practices or amalgamations of firms that would harm society through the exercise of market power.

Some analysts, however, argue that the actual objective of these laws is not efficiency, and that these laws were passed to help certain groups and harm others. For example, some argue that the antitrust laws are designed to help small firms that compete with large firms, whether or not efficiency is increased. In particular, the antitrust laws against price discrimination were passed in response to political lobbying by many small firms that were complaining of larger firms' ability to secure lower prices in their purchases of supplies (Ross 1984).

A group of firms that obtains a general exemption from the antitrust laws can reduce competition and thereby benefit. Many groups have succeeded in obtaining exemptions from the antitrust laws. Workers who unionize in order to raise their wages are specifically exempted from the antitrust laws, as are certain agricultural groups and export associations. Although certain regulated industries such as insurance have obtained antitrust exemptions, regulated industries are generally subject to the antitrust laws. Moreover, as Chapter 20 shows, legislators often try to protect certain groups from competition that is legal under the antitrust laws. It is legal for firms to attempt to influence legislation in order to protect themselves from competition and insulate themselves from antitrust liability (but see Example 19.1).⁴

³The FTC can also bring an action in federal court to obtain a preliminary injunction preventing consummation of a merger.

⁴This lobbying is protected by what is called the Noerr-Pennington doctrine: *Eastern Railroad Presidents Conference v. Noerr Motor Freight, Inc.*, 365 U.S. 127 (1961) and *United Mine Workers of America v. Pennington*, 381 U.S. 637 (1965). We cite cases primarily from the U.S. Reporter (U.S.), Federal Reporter (F.2d), and the Supreme Court Reporter (S. Ct.), which are standard legal references. For example, the Pennington citation appears in vol. 381 of the U.S. Reporter starting at page 637, and that case was decided by the Supreme Court in 1965. A case is first decided in a District Court. It can then be appealed to the Court of Appeals in the relevant region (called a Circuit Court) and after that to the Supreme Court.

EXAMPLE 19.1

Using the Government to Create Market Power: Misuse of the Orange Book

Under the Noerr-Pennington doctrine, firms have the right to petition the government for legislation and then to take advantage of that legislation. That is, a firm has the legal right to lobby for legislation that will make it difficult for others to compete with it, even though such legislation will harm consumers. Recently, the Federal Trade Commission (FTC) has succeeded in reining in firms' ability to misuse government might to create market power.

The FTC has paid particular attention to misuses related to drug approvals. If a firm wishes to make a generic version of a branded drug, it can submit an abbreviated application for a new drug and rely on previous testing of the branded drug in order to obtain approval from the Food and Drug Administration (FDA). The FDA asks makers of branded drugs to list in its "Orange Book" any patents that still apply to their drugs. The FDA will grant an automatic 30-month delay to the introduction of a generic drug that the manufacturer of a branded drug claims violates a patent listed in the "Orange Book." The FDA does not investigate whether the patents listed in its Orange Book are in fact valid. Therefore, it is possible for the manufacturer of a branded drug to list a patent that is invalid or inapplicable and to use it to delay the introduction of a generic rival.

The FTC claimed that Bristol-Myers misused FDA procedures to prevent competition from generic rivals for three of its best-selling drugs, with the result that consumers for these anxiety-preventing and cancer treatment drugs were deprived of generic competition that commonly drives down drug prices by more than 50 percent. The FTC claimed that, among other things, Bristol-Myers filed a false patent to block entry, and that it acquired another patent for the purpose of preventing generic entry. The combined sales of the drugs that Bristol-Myers sought to protect exceeded \$1.5 billion in the year before generics entered. The FTC and Bristol-Myers settled the case with Bristol-Myers agreeing to restrictions on its conduct.

Source: John Wilke, "Bristol-Myers Settles Patent-Law Abuse," *Wall Street Journal*, March 10, 2003.

The view that the guiding principle of the antitrust laws should be efficiency, rather than the taking of resources from one group and granting them to another, has gained increasing acceptance among legal and academic scholars. One appeal of such a simple proposition is that it provides a clearer guide as to what antitrust policy should be than does the alternative view of helping "deserving" groups.

Even if one accepts the proposition that the goal of the antitrust laws is to promote efficiency, economists often have difficulty determining which practices result in inefficient behavior. For example, suppose that two firms merge and the resulting reduction in competition causes price to rise. That sounds bad. However, suppose that, as a

result of the merger, the merged firm develops a new and better product or provides the same product but offers better services or develops a lower-cost method of production than before. That sounds good. Should the antitrust laws ban all mergers if they significantly eliminate competition, or should they also pay attention to the potential efficiency gains and balance the two?

To see how the trade-off between an increased price and increased efficiency in other dimensions can be compared, suppose that, as a result of a merger, a firm raises its price from \$1 to \$1.10 because of the elimination of competition, which causes a deadweight loss (a triangle in Figure 19.1). Suppose that the merger also enables the firm to operate more efficiently and lower its constant marginal cost from \$1 to \$0.90, which results in a greater productive efficiency (a rectangle in Figure 19.1).

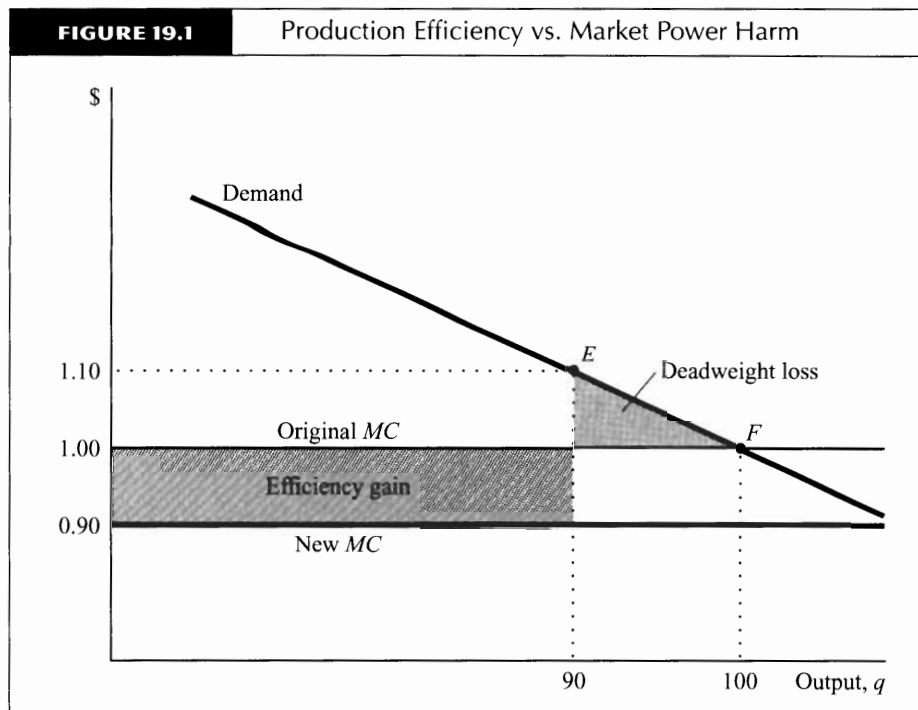
If the triangular area representing the deadweight loss from the price increase is smaller than the rectangular area of efficiency gain, the merger is, on balance, good for society. The relative size of these two areas depends on the particular circumstances. The larger the quantity sold in the marketplace, the more important the efficiency gains, and the larger the area of the rectangle compared to the triangle. Even small reductions in costs per unit can result in efficiency gains that swamp deadweight loss in importance.⁵

For example, suppose that the initial quantity is 100 units (initial equilibrium point F), and the postmerger quantity is 90 (equilibrium point E). Because the efficiency savings are 10¢ per unit, the efficiency gains are \$9. The deadweight loss from the price increase is approximately 50¢ ($= -1/2 \times 10¢ \times [-10] \cong -1/2 \Delta p \Delta Q$). Thus, the efficiency gains outweigh the deadweight loss.

These types of calculations can be complicated, and it is a matter of debate whether courts should be charged with making such calculations in deciding the legality of a merger (Williamson 1968–69). The current policy statement of the Department of Justice and the Federal Trade Commission, the *Horizontal Merger Guidelines*, explicitly recognizes the importance of efficiency gains in evaluating mergers. However, these guidelines suggest in general that a merger would be challenged if it has an anticompetitive effect (a price increase) even if there is an offsetting efficiency gain.

Most other antitrust authorities also forbid mergers that raise price even if total efficiency (producer surplus plus consumer surplus) rises. Two notable exceptions are Australia and New Zealand, countries that rely heavily on international trade and for which efficient export industries are a key to economic prosperity. Lyons (2002) shows how a merger standard that prohibits efficiency-enhancing mergers that increase price can lead to greater efficiency than one that allows mergers as long as the merger increases efficiency. The reason is that if a firm is prevented from undertaking one efficiency-enhancing merger (a merger in which price rises), it may engage in another

⁵An efficiency gain, which depends on the total quantity produced, tends to swamp the deadweight loss, which depends on the restriction in output resulting from the merger if the output restriction is a small percentage of the total quantity produced. Notice that Figure 19.1 is based on the assumption that the price is initially set competitively. Were the price initially set higher, the deadweight loss would be a trapezoid comprising a triangle plus a rectangle whose width is 10 units and its height equals the initial gap between price and marginal cost.



merger that ultimately creates greater efficiency for society (even though profit to the merging firm will be lower).

Requiring courts to apply sophisticated economic analyses to evaluate behavior may be unrealistic. Moreover, the courts must often deal with economic problems that economists have not yet analyzed. Courts don't have the luxury of taking as much time as necessary to solve a problem. Still, it is hard to argue that economic knowledge should be ignored. Use of economics in antitrust proceedings has increased worldwide. As a result, litigation is more complicated because economic analysis can now be the centerpiece of a case, and different countries can have conflicting analyses. (See Example 19.2.)

Who May Sue?

Individuals and firms, in addition to the Department of Justice and the FTC, can bring antitrust suits. Deciding who has **legal standing**, which is the right to bring a suit, is complicated. Only a party that suffers an injury that the antitrust laws were designed to prevent is permitted to sue. For example, suppose that if two firms merge, they will become very efficient and reduce price. Rival firms would be harmed by the merger, yet they have no legal standing to sue to block the merger under the antitrust laws because the goal of the antitrust laws is to generate low prices to consumers.⁶

⁶*Brunswick Corp. v. Pueblo Bowl-O-Mat, Inc.*, 429 U.S. 477 (1977) and *Cargill Inc. v. Monfort of Colorado*, 197 S. Ct. 484 (1986).

EXAMPLE 19.2*Conflict Between European and U.S. Antitrust Authorities: GE-Honeywell*

In October 2000, General Electric (GE) announced plans to merge with Honeywell. GE manufactures many products, including aircraft engines for commercial aircraft. Honeywell does not produce aircraft engines, but does make products used in aviation such as weather radar, engine accessories, braking equipment, aircraft lighting, and environmental control systems. Thus, the merger would not involve both firms producing the same product or raise vertical supply relations issues because the merging firms manufactured only complementary products. How then could this conglomerate merger present an antitrust issue?

As we saw in Chapter 11, when a single firm prices two complementary products, it takes into account that lowering the price on one product will stimulate demand for the other. This demand effect creates an incentive for the firm to set lower prices for both products than would be set if each product were sold by a different firm. Thus, one should expect lower prices and consumer benefits from the GE-Honeywell merger due to more efficient pricing. Using this reasoning, U.S. antitrust authorities did not challenge the merger.

Because GE and Honeywell have large worldwide sales, their merger also fell under the scrutiny of European antitrust authorities. European authorities disagreed with U.S. authorities and prevented the merger in July 2001. The European authorities reasoned that, because the combined firm would have an incentive to lower prices, it would “dominate” the market, and hence the merger should be prevented. This convoluted reasoning is a throwback to earlier U.S. antitrust history when regulators viewed the creation of large, efficient firms as undesirable because it would drive less efficient rivals out of business even though the greater efficiency would benefit consumers. Most economists would argue that the protection of competition, not competitors, should be the focus of antitrust law.

This failed merger and the conflict among antitrust authorities has spawned efforts to harmonize antitrust enforcement and standards. Such harmonization is one of the goals of the International Competition Network, a recently formed organization comprising all major antitrust authorities. Preventing foreign antitrust authorities from repeating past mistakes of U.S. antitrust policy would be a major contribution of this organization.

Source: Nalebuff (2004) in Kwoka and White (2004).

Suppose that several manufacturing firms engage in a conspiracy to raise prices and that they sell their products to a retail department store, which then sells them to final consumers. The retail department store is not engaged in any conspiracy and simply applies the usual retail markup to the price it is charged. Who should be allowed to sue the manufacturers who are engaged in the conspiracy? The department stores, which

are direct purchasers from the conspiring parties, certainly have the right to sue. Do final consumers (indirect purchasers) have a right to sue? It would be double counting to allow both consumers and retailers to collect damages from the same overcharge. In *Illinois Brick*, the Court limited the right of indirect purchasers to sue.⁷ One possible justification for this decision is the difficulty of determining all of the potential indirect and direct parties who would be entitled to sue and to obtain duplicative damages for the same price overcharge from the plaintiffs (Landes and Posner 1979).

Suppose that individual consumers are the direct purchasers of a product from firms that have engaged in a conspiracy. Would an individual purchaser have the incentive to sue if all that could be recovered was the price overcharge on the product, trebled, plus attorneys' fees? Usually no. In an effort to create greater incentives to bring suits, courts allow attorneys to file class-action suits on behalf of all consumers. Attorneys are entitled to legal fees, and this provides an incentive for them to bring such suits. Of course, attorneys have an incentive to bring too many such cases if they are overcompensated.

Economic Theory of Damages

Economists use their theories to recommend the optimal amount of damages that a guilty defendant should pay. Although economic theory is widely used to determine antitrust liability, it is less commonly used to determine optimal damages.

The economic theory of damages starts from the proposition that the purpose of damages is to deter inefficient activity but not to be so burdensome as to deter efficient activity. For example, suppose that any time a firm was convicted of illegally conspiring with another firm, the firms' managers were executed. As will be shown, it is not so easy to determine when an agreement among firms is an illegal one. If the penalty were death (or otherwise extremely harsh), many firms might be dissuaded from activity that could be perfectly lawful and beneficial, such as the creation of a trade association that sets product safety standards. The optimal penalty is one that balances the beneficial and the nonbeneficial aspects of deterrence.

An optimal penalty reduces the incentive to engage in illegal activity. For example, suppose that, if a group of firms conspire, they can raise their profits by \$100. If this conspiracy could be detected with certainty and at no cost by enforcement officials, then a penalty of \$100 would suffice to deter the activity. Of course, deterrence is not perfect, and it takes resources to discover illegal activity. Because the firms know that they may not be caught, a penalty higher than \$100 may be necessary to deter cartel behavior. For example, suppose that price-fixing conspiracies are detected in only one-third of all cases. Then, as a first approximation, charging a penalty of \$300 should be enough to deter the illegal activity.⁸

An additional complication arises with an international cartel. If the purpose of a damages award is to deter, then the damages should deprive the cartel of its full profits.

⁷*Illinois Brick Company v. Illinois*, 431 U.S. 720 (1977). The word *court* spelled with a capital C refers to the Supreme Court of the United States.

⁸For more on optimal damages in antitrust cases, see Landes (1983) and White (1989). According to Landes, the optimal penalty equals the harm the firms impose on others, adjusted for the probability of detection.

But if some foreign country does not have an antitrust law, the cartel can earn excessive profits in that country unless foreign consumers can sue for damages in the United States or other countries with antitrust laws. The possibility of bringing suit in another country raises complicated jurisdictional issues that remain unsettled.⁹

Suppose that a group of firms is convicted of violating the antitrust laws. How should the penalties be apportioned among the firms? Should one defendant pay all of the damages, or should there be some sharing rule? In particular, suppose that of the two defendants to a lawsuit, one defendant settles with the plaintiff (for example, pays the plaintiff \$100 to be dismissed from the lawsuit) before the case goes to trial. At the trial, the damage award to the plaintiff is several times the amount of the settlement. Should the remaining defendant be forced to pay the entire amount? The Court has ruled that an antitrust defendant is not entitled to **contribution**, which is a payment to a guilty defendant from other culpable parties.¹⁰ This decision has been criticized on the grounds that it appears unfair. However, the incentive to settle will be greatest if, once the settlement is achieved, no further liability can be assumed, and therefore a large damage award can be avoided. Thus, this rule greatly increases the incentive to settle, with a resulting savings in litigation costs (Easterbrook, Landes, and Posner 1980).

The Use of U.S. Antitrust Laws

U.S. antitrust laws have been used increasingly to prosecute price-fixers, and penalties have increased.¹¹ From 1890 through 1974, the Justice Department brought 1,000 civil and 723 criminal cases (Posner 1976, 25). The penalty in a criminal case can be a jail sentence, but the penalty in a civil case cannot. Since World War II, Democratic and Republican administrations have brought cases at roughly the same rate (Posner 1970, 411–12). The Department of Justice wins most of the cases it brings. In every five-year period since 1910 studied by Posner (1970, 381–82), the Department won at least 64% of all its cases; since 1925 it has won at least 78%; and since 1955 it has won at least 85%.¹²

The Department of Justice loses few criminal cases (Posner 1970). Since 1890, 57% of the cases have been disposed of on *nolo contendere* (“no contest”) pleas, 21% on other convictions, and only 22% resulted in acquittals and dismissals. The recent

⁹See *S.A. Emparagan, et al. v. F. Hoffman-La Roche, Ltd. et al.*, 315 F.3d 338 (D.C. Circuit, 2003). The case is before the Supreme Court in 2004.

¹⁰*Texas Industries, Inc. v. Radcliff Materials, Inc.*, 451 U.S. 630 (1981).

¹¹Cases brought by states under their own antitrust laws were extremely infrequent prior to the 1970s. In recent years, a few states have made increasing use of their antitrust laws, but state cases are still uncommon.

¹²Posner (1970, 384) also shows that the Department of Justice has won 74% of its cases before the Supreme Court, the Federal Trade Commission has won 75%, and private claimants have won 63%. Averaged across the three groups, 70% of these cases before the Supreme Court were won by the plaintiffs.

conviction rates have been even higher (Snyder 1990). During the early 1990s, the conviction rates exceeded 90%. These high success rates may indicate that the Department only tries sure cases or that defendants plead *nolo contendere* and pay small fines (avoiding the costs of long court battles and the risks of losing). One advantage of a *nolo contendere* plea for a defendant in a government case is that it does not provide evidence of collusion that can be used subsequently in a private antitrust action seeking treble damages.

The penalties imposed by the federal antitrust laws have historically been relatively small. The FTC's only remedy is an injunction called a *cease-and-desist order*, which prohibits the behavior but does not penalize the firm for engaging in it. The maximum fine under the original Sherman Act was \$5,000 and the maximum prison sentence was one year. The maximum fines were increased to \$50,000 in 1955. In 1974, they were increased to \$100,000 for individuals and \$1 million for corporations. In the 1950–1959 period, the fines in the Justice Department's cases averaged \$40,000 (0.08% of the sales involved in the conspiracy), whereas from 1960 to 1969 they averaged \$131,000 or 0.21% of the sales involved in the conspiracy (Posner 1970, 1976). Gallo et al. (1994) estimate that fines averaged 1% of the present value of conspiracy sales from 1985 to 1993. The maximum fines are currently \$10 million for a corporation and \$350,000 for an individual. However, the fines can be increased up to twice the violator's gain or twice the victims' loss (Berkman 1997). Beginning in 1997, fines levied on individual firms and total fines across all firms increased enormously. The Department of Justice collected total criminal fines of less than \$50 million per year between 1990 and 1996. However, from 1997 to 1999, total fines rose from \$205 million to over \$900 million per year, but they declined to \$102.5 million in 2002, and to \$64.2 million in 2003. The largest fines on individual firms were \$500 million for Hoffman-La Roche and \$225 million for BASF (1999 vitamin price-fixing case); \$135 million for SGL Carbon, \$134 million for Mitsubishi, and \$110 million for UCAR Carbon (1999, 2001, and 1998 cases, respectively, concerning the fixing of the price of graphite electrodes used in steel furnaces); and \$100 million for Archer Daniels Midland (1997 lysine and citric acid price-fixing case).¹³ In addition to fines, a guilty firm must recompense private individuals, firms, and government bodies who prove in court that they were damaged from the price-fixing conspiracy. Individuals and firms are entitled to treble damages plus attorneys' fees, and the U.S. government receives damages plus attorneys' fees.

Prison sentences are rare in federal antitrust cases. There were none during the period from 1890 to 1909. From 1910 to 1974, there were 33 cases in which a prison sentence was imposed, roughly one every other year (Posner 1976, 33). Prior to 1925, most of those jailed were union organizers and officers. Until after World War II, most prison sentences for price fixing were restricted to cases involving violence. The maximum jail sentence was increased to three years in 1974. The average prison term over the period from 1955 to 1993 was about three months (Gallo et al. 1994).

¹³U.S. Department of Justice, "Sherman Act Violations Yielding a Fine of \$10 Million or More," January 23, 2003: www.usdoj.gov/atr/public/criminal/202532.htm.

In 2002, of the 36 individuals sentenced in antitrust cases, 19 were sentenced to incarceration time for a cumulative total of 10,501 days (16 individuals were sentenced to other confinement such as house arrest or a halfway house for a total of 3,607 days).

Private Litigation

Private actions brought by individuals or firms harmed by anticompetitive behavior were rare for price-fixing cases until the electrical conspiracy of the early 1960s (see Example 5.1). In the next two decades, private actions, including class-action suits, increased substantially. These private actions, which often follow federal suits, can substantially increase the cost of conspiracy, because treble damages plus attorneys' fees may be collected.¹⁴ For the period from 1937 to 1954, there was an average of 104 private cases per year; for the period from 1955 to 1959, the average increased to 229 cases per year. From 1960 to 1964, 1,919 cases were filed concerning electrical equipment, causing the yearly average to rise to 671 cases. For fiscal years ending June 30th, the number of private civil antitrust suits fell from 1,457 in 1980 to 1,100 in 1984 to 638 in 1989 (Salop and White 1986, Abere 1991).

Market Power and the Definition of Markets

The antitrust laws concentrate on controlling the creation and maintenance of market power. This section defines market power, discusses how to measure it, and points out that it is sometimes difficult to measure market power accurately. Some economists and lawyers argue that one should define a market and then calculate the market share of the firm under analysis as an approximation of its market power, where a high market share is interpreted as an indicator of market power. This section describes the principles that should be used to define a market and provides a background against which cases can be interpreted.

Market Power

A firm (or group of firms acting together) has market power if it is profitably able to charge a price above that which would prevail under competition, which is usually taken to be marginal cost. This ability to set price above marginal cost implicitly uses the model of perfect competition as a benchmark against which to measure the behavior of firms. If this definition is applied literally, probably every firm in the United States has at least a tiny bit of market power. The model of perfect competition is an extreme one that describes few, if any, actual industries. Therefore, presumably, when courts find that a firm has market power, they must mean the firm has a substantial

¹⁴According to Posner (1970, 372), in the period from 1956 to 1960, at least 278 private cases were preceded by a Department of Justice judgment, while in the period from 1961 to 1963, 880 cases followed such judgments.

amount of market power for some significant period of time. Unfortunately, the courts have not stated how much market power is needed. Does a price 5 percent above marginal cost for two years reflect substantial market power? Or 10 percent above for one year?

It is difficult to measure marginal cost and therefore difficult to measure the deviation between price and marginal cost, even if the courts stated how substantial a deviation must be to constitute significant market power. An alternative approach is to estimate the price elasticity of the residual demand (the market demand net of the quantity supplied by other firms) facing an individual firm (or group of firms). This elasticity of residual demand facing a firm summarizes the ability of a firm (or group of firms acting together) to exercise market power. The price-cost margin equals the negative of the inverse of the elasticity of demand (Chapter 4): $(p - MC)/p = -1/\epsilon$, where p is price, MC is marginal cost, and ϵ is the elasticity of residual demand.¹⁵

If the elasticity is large, the firm has little market power. Most empirical estimates of demand curves for individual firms selling branded products rarely find the absolute value of price elasticities to be higher than 5 to 10 (Telser 1972, 274–306). Using the price-cost margin formula, if the elasticity is -5 , price is 25 percent above marginal cost; if the elasticity is -10 , the price is 11 percent above marginal cost.

Whether a firm currently has market power is a much different question from whether, as a result of a merger, it could acquire and exercise *additional* market power. The first question, whether price is already elevated significantly above competitive levels, can be answered directly by comparing price and marginal cost or indirectly by looking at the elasticity of demand facing the firm. The second question, whether price will rise significantly above its current level as a result of the merger activity, can be answered directly by predicting how price will change or indirectly by predicting how the elasticity of demand facing the firm will change as a result of the merger.

For example, suppose a firm makes a differentiated product, Product A. The residual demand curve for A can be estimated as a function of the price of Product A and the price of another firm's substitute product, Product B. If the direct elasticity of A at current prices is very large, then the firm producing A has no market power. In a merger case, economists attempt to predict how the elasticity of each product will change as a result of the merger. For example, suppose the firms producing A and B want to merge. The merger between the firms will allow the merged firm to set prices jointly, and the analyst can calculate the resulting prices based on the demand curve facing the merged firm (Baker and Bresnahan 1985, Hausman et al. 1994). The

¹⁵In a static model, the elasticity of residual demand completely summarizes a firm's market power and determines its pricing. In a dynamic model, this simple relationship does not hold because the profit-maximizing price depends on the responses of consumers and firms over time. Moreover, in dynamic oligopoly models, the equilibrium depends on the game played by rivals. Another approach to measuring market power involves calculating the long-run rate of return. See Chapter 8 for a discussion of the caveats of using this approach.

merger increases market power if the postmerger prices are higher than the premerger ones.¹⁶

Sometimes economists cannot estimate a price elasticity accurately because the data are inadequate or unavailable. In an attempt to reach some workable solution to the problem of determining market power, analysts and the courts often define a market and then construct a measure of market share. If the market share of the firm (or firms) under analysis is high, the suggestion is that market power exists. In a merger case, the Justice Department or the FTC looks at whether there will be a significant increase in concentration as a result of the merger. There is no agreement as to exactly what share (or change in share) is “high,” but many economists regard a share in the range of 30 to 50 percent as too low to indicate significant market power in an industry with a competitive fringe comprising the remainder of the market.

Market shares are imperfect indicators of market power, so additional analysis of the economic conditions is necessary before one can reach a conclusion about market power. For example, if entry is easy, then the industry pricing is severely constrained regardless of whether an existing firm has a large market share. Similarly, the presence of factors that make it difficult to maintain a cartel is relevant (Chapter 5).

Market Definition

In merger or other antitrust cases, economists are often called on to define a market. Were it not for these cases, it is doubtful whether such a large body of economic research would have developed on defining markets.

Alfred Marshall (1920, 324) defined a market as an area in which “prices of the same goods tend to equality with due allowance for transportation costs.” Since Marshall’s time, economists and lawyers have refined the definition of a market. A **market definition** specifies the competing products and geographic area in which competition occurs that determines the price for a given product. Clearly Coke and Pepsi are in the same market. But is Dr Pepper or Canada Dry Ginger Ale in that market? Is milk?

How the market is defined often determines the outcome of antitrust cases. For example, in determining whether to permit a merger, the government and the courts examine the market shares of firms, which are viewed as proxies for the firms’ actual or potential market power. A firm’s market share depends crucially on the market definition. Coke’s share of its market will be much larger if the market is defined as colas than if it is defined as all soft drinks or all drinks. Example 19.3 discusses how the government applies the principles of market definition in its antitrust enforcement policy toward mergers.

¹⁶A merger simulation typically estimates demand curves; assumes that firms engage in a particular game (such as Bertrand); calculates marginal cost by using observed prices and estimated elasticities; and predicts new prices assuming that the merged firm coordinates pricing among the products it controls. See Example 7.5 and see Carlton (2003, 2004a) for a discussion and critique of this approach.

EXAMPLE 19.3*The Merger Guidelines*

The government policy on the definition of markets is contained in the *Horizontal Merger Guidelines* that the Department of Justice and the Federal Trade Commission (FTC) issued in 1992. These *Merger Guidelines* set forth the principles that the government uses to define relevant economic markets in its enforcement of the antitrust laws regarding merger activity. These guidelines were slightly revised in 1997.

According to the *Merger Guidelines*, the government's first step in reviewing a proposed merger is to properly define the market. The *Merger Guidelines* specify that a market is the smallest group of products and the smallest geographical area such that a hypothetical monopoly of all those products in the area could raise price by a certain amount (for example, 5 or 10 percent) above any prevailing or likely future levels. One problem with this definition is that according to the definition, a small group of firms could constitute a market even though fringe firms existed that produced the identical product. Stigler and Sherwin (1985) raise other objections as well.

After defining the market, the *Merger Guidelines* require that the government determine whether the proposed merger will greatly increase concentration (and hence, presumably, market power). Concentration is measured using a Herfindahl-Hirschman Index (HHI), which is the sum of the squared market shares (expressed as percentages) of firms in the industry. The *Merger Guidelines* do not regard a merger between two firms as raising concerns about competition if the postmerger HHI in the industry is below 1000. If the postmerger HHI is between 1000 and 1800, the merger does raise concerns if the change in the HHI resulting from the merger is 100 points or more. If the industry's postmerger HHI is over 1800, the merger raises concerns about competition if the change in the HHI resulting from the merger is 50 points or more.

For example, suppose that an industry consists of four firms, each with a 25 percent market share, and two of the firms want to merge. The initial HHI is $2500 (= 25^2 + 25^2 + 25^2 + 25^2)$, and the postmerger HHI is $3750 (= 50^2 + 25^2 + 25^2)$. Because the postmerger HHI exceeds 1800 and the change in the HHI exceeds 50, concerns about competition are raised according to the *Merger Guidelines*.

The application of the *Merger Guidelines* implicitly assumes that after a merger, the firms involved will be able to maintain their premerger market shares and the merged firm will enjoy a market share equal to the sum of the premerger shares of the firms that are merging. When this assumption is not reasonable, the analysis should be modified in order to properly reflect the merged firms' market share. The *Merger Guidelines* recognize that other factors (such as ease of entry) in addition to market concentration influence market behavior. Both the Department of Justice and the FTC consider those factors before deciding to challenge a merger.

The Extent of the Product Market. A proper definition of the product dimension of a market should include all those products that are close demand or supply substitutes.¹⁷ Product B is a *demand substitute* for A if an increase in the price of A causes consumers to use more B instead. Product B is a *supply substitute* for A if, in response to an increase in the price of A, firms that are producing B switch some of their production facilities to the production of A.¹⁸ In both cases, the presence of B significantly constrains the pricing of A, provided that an increase in the price of A would result in either a significant decline in the quantity of A consumed as consumers switch from A to B or a significant increase in the supply of A as firms switch production from B to A.

The degree of substitution between products depends on the current prices of the two products. For example, A and B may be highly substitutable at a high price for A, but not at a low price for A. Even a monopoly may raise its price sufficiently above competitive levels so that eventually it faces some competition from other products. Just because a monopolized product faces close demand substitutes at the monopoly price, it does *not* follow that the firm producing the product has no market power (though it may not be able to raise its price further). It is only if the substitution possibilities are so large as to generate a highly elastic residual demand that the monopoly has no significant market power. Because it is difficult to determine which products to include in the market definition, market shares may be only a crude indicator of market power.

The *Cellophane* case illustrates these difficulties in defining a market.¹⁹ The Court investigated whether du Pont had market power in the pricing of cellophane. The Court reasoned that du Pont lacked market power because, at the current market prices, a user of cellophane had many substitutes, such as paper bags, and du Pont's share of the market including these substitutes was not large. There was also evidence, however, that price substantially exceeded marginal cost. Based on the foregoing discussion, it was an error to include other wrapping materials in the market definition because they did not prevent the exercise of market power and constrain the price of

¹⁷The relevant economic market is not necessarily the same as the *market* that a salesperson might refer to. Substantial confusion has sometimes resulted when market definition is based on memos written by marketing personnel. For that reason, some antitrust lawyers advise companies to instruct marketing personnel to avoid the use of the word *market* in memos.

¹⁸The relationship between the demand elasticity facing a firm and supply and demand substitutes can be illustrated using the model of a dominant firm facing a competitive fringe (see www.aw-bc.com/carlton_perloff "Dominant Firm and Competitive Fringe Model" and Landes and Posner 1981). One can derive that

$$\epsilon_d = \frac{Q}{Q_d} \epsilon - \frac{Q_f}{Q_d} \eta_f,$$

where ϵ_d is the residual demand elasticity facing the dominant firm, ϵ is the market demand elasticity, Q is the market quantity, Q_d is the quantity sold by the dominant firm, Q_f is the quantity supplied by the fringe, and η_f is the supply elasticity of the fringe. As the absolute value of ϵ increases (more demand substitutes) and as η_f increases (more supply substitutes), and as the dominant firm's share (Q_d/Q) falls, ϵ_d rises in absolute value and the dominant firm has less market power.

¹⁹*United States v. E.I. du Pont de Nemours & Co.*, 351 U.S. 377 (1956).

cellophane to competitive levels. If, however, instead of asking whether du Pont had market power, the Court had investigated whether a proposed merger would raise the cellophane price, its market definition might have been appropriate.

In the *Brown Shoe* case, the Supreme Court articulated a laundry list of criteria that can be used to define markets.²⁰ It said: “The boundaries of such a submarket may be determined by examining such practical indicia as industry or public recognition of the submarket as a separate economic entity, the product’s peculiar characteristics and uses, unique production facilities, distinct customers, distinct prices, sensitivity to price changes, and specialized vendors.” The application of this laundry list of criteria has not led to precision in defining a market for antitrust purposes.²¹

Numerous methods are used to identify the good substitutes for a particular product. One is to interview producers in the industry who presumably know both their customers and their potential competitors from other industries.

If Products A and B are in the same economic market, then their prices should tend to move closely together. Therefore, a reasonable *first* step in defining economic markets is to examine the price correlations (a statistical measure of how closely prices move together) among different products that are under consideration for inclusion in the same product market.²²

Although no standard levels of correlation have been established to determine if two products are in the same market, the available data may often be used to develop such standards. For example, suppose that everybody agrees that two different types of plastic materials are in the same economic market. One could compute the correlation between their prices and use it as a benchmark to determine whether some third plastic material belongs in the same economic market with the other two products.

The direct price elasticity—*not* the cross-elasticity of demand—determines market power. The **cross-elasticity of demand** is the percentage change in quantity demanded in response to a 1 percent change in the price of another product. There is a lot of discussion in court decisions as to the importance of cross-elasticity of demand in defining markets. Courts often use the term loosely to indicate that products are substitutes. There is a relationship between cross-elasticity and direct elasticity,

²⁰*Brown Shoe Company v. United States*, 370 U.S. 294 (1962).

²¹The courts, in addition to defining economic markets, have occasionally attempted to define economic *submarkets* that are contained within an economic market. Presumably, competition between two products in the same economic market is more intense if the two products also belong to the same submarket. The distinction between *market* and *submarket* is not very useful, and we will not refer to it or even attempt to give an economic definition of the term *submarket*.

²²Price correlations are a useful first step in defining markets; however, high correlations need not always indicate that two products are in the same market. For example, dissimilar products made from similar inputs may have high price correlations. Similarly, low correlations need not always indicate that products are not in the same market, provided large quantity shifts accompany the relative price shifts. If the price of one product rises, but the price of a good substitute does not, the quantity demanded of the first product sharply declines.

however. All else the same, the larger a cross-elasticity of demand, the larger in absolute value is the direct elasticity of demand.²³

To intelligently discuss a cross-elasticity, one must specify whether it is the cross-elasticity of Product A with respect to the price of Product B or vice versa. Although these two different cross-elasticities are usually not distinguished in court decisions, they are not equal in general.²⁴ The relevant cross-elasticity of demand when the question is whether the market for Product A should include Product B is the cross-elasticity of demand for Product A with respect to the price of Product B.

The Extent of the Geographic Market. The geographic limit of a market is determined by answering the question of whether an increase in price in one location substantially affects the price in another. If so, then both locations are in the same market. The process of determining these limits proceeds along the same lines as discussed for the product market definition and involves similar reasoning. For example, consider the consumption of oranges in Chicago. Oranges are shipped to Chicago from outside the city limits. The geographic areas that ship to Chicago (or could profitably do so if price rose slightly) are in the same economic market as Chicago because they contain orange producers whose output significantly influences the price of oranges in Chicago. Notice that these same orange producers could also significantly affect the price of oranges in Milwaukee. Thus, Milwaukee and Chicago could be in the same economic market, and the price of oranges in Chicago would generally be closely related to the price of oranges in Milwaukee.²⁵

Cooperation Among Competitors

This section explores the restrictions that the antitrust laws place on cooperation among competitors. We first examine explicit agreements to set price or output and then explicit agreements that lead to new products being produced, as well as informa-

²³This result follows because the sum of the direct elasticity plus all cross-elasticities of demand equals 0. Let the cross-elasticity of demand of Product A with respect to the price of B be $\epsilon_{AB} = (\partial Q_A / \partial p_B)(p_B / Q_A)$, where Q_A is the (income-compensated) demand for A, and p_B is the price of B. Then, $0 = \epsilon_{AA} + \sum_B \epsilon_{AB}$, where ϵ_{AA} is the own (direct) price elasticity of demand for product A (Henderson and Quandt 1980, 31–33). The cross-elasticity of demand is positive for substitutes, and the direct price elasticity is negative. The direct elasticity can be large even if no individual cross-elasticity is large.

²⁴From demand theory, $\partial Q_A / \partial p_B = \partial Q_B / \partial p_A$. This last relationship does not imply that the cross-elasticities of demand (defined in the previous footnote) ϵ_{AB} and ϵ_{BA} are equal (Henderson and Quandt 1980, 30).

²⁵See Carlton (2003, 2004a), Landes and Posner (1981), Scheffman and Spiller (1987), and Stigler and Sherwin (1985) for further analysis of market definition and its use in antitrust cases.

tion sharing among rivals. We then analyze oligopoly behavior in which firms behave similarly, though not as a result of explicit agreements. Finally, we examine mergers among competitors.

Price-Fixing and Output Agreements

The Court's views on price-fixing and output agreements are that an agreement whose sole purpose is to eliminate competition and raise prices above competitive levels—that is, a “naked” agreement to eliminate competition—is illegal. No inquiry as to the reasonableness of the price set is necessary to reach the conclusion that the agreement violates the law. When no additional inquiry is necessary to analyze the facts of a situation in order to determine the legality of the conduct, the conduct is said to be *per se* illegal. Therefore, it is often said that price-fixing and output-fixing agreements are *per se* violations of the antitrust laws.²⁶ Example 19.4 discusses the approach that other countries have taken toward such agreements among competitors.

Soon after the passage of the Sherman Act, the courts considered two cases of firms that cooperatively set prices and allocated customers. In *Trans-Missouri Freight Association*, a group of competing railroads entered into agreements about what rates to charge.²⁷ The railroads claimed that the rate agreements resulted in reasonable rates that prevented ruinous competition. The Court rejected these arguments and instead ruled that “the claim that the Company has the right to charge reasonable rates and that therefore it has the right to enter into a combination with competing roads to maintain such rates cannot be admitted. . . . Competition will itself bring charges down to what may be reasonable. . . .”

After the decision limiting their ability to fix rates, railroads continued to push for the ability to set their rates and avoid competition. They were ultimately allowed to do so by legislation. The Staggers Act of 1980 eliminated many of the restrictions on competition. Apparently as a result, many railroad mergers followed.

Considered at almost the same time as *Trans-Missouri*, the *Addyston Pipe* case also involved price fixing.²⁸ A group of manufacturers of cast-iron pipe met to set price terms in certain geographic areas. Their defense was that the prices they set were fair and reasonable and restrained the deleterious effects of ruinous competition. The

²⁶It is ironic that an agreement among two small competitors in a very competitive industry where the two together cannot affect the market price is a *per se* violation of the antitrust laws while a merger of the two firms is legal. If agreements among competitors can never generate efficiencies, then it lowers enforcement costs to ban both effective and ineffective agreements to fix price among competitors. However, just like mergers, agreements among competitors can generate efficiencies, so it is peculiar to distinguish between the two, unless one could claim that efficiencies are more likely from a merger than from a price-fixing agreement. As we discussed in Chapter 5, it is often unclear what the word *agreement* means. Here, we refer to explicit communication among firms about what specific price to charge or output to produce. See Carlton, Gertner, and Rosenfield (1997).

²⁷*United States v. Trans-Missouri Freight Association*, 166 U.S. 290 (1897).

²⁸*United States v. Addyston Pipe & Steel Co.*, 175 U.S. 211 (6th Cir. 1899).

EXAMPLE 19.4*Antitrust Laws in Other Countries*

Many countries took a much different view toward agreements among competitors than that taken in the United States. For example, Germany, Japan, and the United Kingdom allowed the formation of cartels that the government believed would promote efficiency. Although competition was generally viewed as desirable, these countries also believed that in certain circumstances it would not lead to efficiency. For example, West Germany and Japan both allowed for cartelization so that firms could reduce capacity during periods of excess capacity. Audretsch (1987) showed that in West Germany, prices rose during the cartelization period and fell after the cartel dissolved.

In addition to having their own laws, countries belonging to the European Union (EU) are also governed by antitrust provisions under the Treaty of Rome as well as by the European Commission Merger Regulation. The main antitrust provisions under the Treaty of Rome are Articles 85 and 86. Basically, Article 85 prevents anticompetitive harms resulting from contracts (such as those that restrict distribution terms) or from agreements (such as cartels). Similarly, Article 86 prevents anticompetitive harm from "abuse of dominant position" (such as from predatory pricing). The EU now uses its antitrust laws to question mergers and its competition policy to prevent price fixing and other similar activities. Thus, the EU laws now more closely resemble U.S. law in enforcing competition. The eastern and central European countries, formerly in the Soviet bloc, are adopting laws similar to those of the European Union.

The International Competition Network, an organization comprised of the antitrust authorities of over 85 countries, was recently formed to promote the development of antitrust laws throughout the world.

Sources: Audretsch (1987) and Swann et al. (1974).

Court reaffirmed its rejection of this type of argument. Several months after the adverse decision, all of the defendants merged into a single firm. The government did not challenge the merger. Therefore, the firms were able to achieve through merger what they were unable to achieve through horizontal agreements. These two early important price-fixing cases both involved industries with high fixed costs that made claims of ruinous competition.²⁹

Nearly three decades later, the Court reinforced the prohibition against price fixing in *Trenton Potteries*.³⁰ The firms that manufactured and distributed 82 percent of the

²⁹It is possible in high-fixed-cost industries for no equilibrium to exist without additional restrictions on competition (see www.aw-bc.com/carlton_perloff "Theory of the Core"). See Bittlingmayer (1982) for an analysis of the *Addyston Pipe* case from this point of view.

³⁰*United States v. Trenton Potteries Co.*, 273 U.S. 392 (1927).

bathroom fixtures produced in the United States attempted to set list prices and urged adherence to these prices. Without investigating whether the agreement could successfully affect prices, the Court ruled that the reasonableness of price was no justification for price fixing.

During the 1930s, there was a widespread belief that the forces of competition were, in large part, the cause of the Great Depression. In 1933, in an apparent response to the current thinking of the time, the Court contradicted its previous rulings on price fixing and ruled in *Appalachian Coals* that a price-fixing cartel could be desirable if it prevented financial ruin. This anomalous decision was overruled by the Supreme Court in 1940. In *Socony-Vacuum*, the Court reaffirmed its previous views regarding price agreements among competitors.³¹ In that case, a group of oil producers formed an organization designed to raise prices in order to rescue the industry from its serious financial plight. The Court ruled that “the elimination of so-called competitive evils is no legal justification” for such programs.

The reduction in the number of cartels is likely the most important achievement of U.S. antitrust laws. Cartels can significantly raise prices to consumers. So, for example, bid rigging conspiracies led to a price increase of about 6.5 percent for milk (Porter and Zona 1999; see also Pesendorfer 2000), 19 percent for highway construction in North Carolina (Brannman and Klein 1992), and 23–30 percent in Defense Department auctions for frozen fish (Froeb, Koyak, and Werden 1993). (See Chapter 5 and Connor 2003 for many additional examples involving international cartels in citric acid, vitamins, and other goods.) Consumers benefited greatly from the prosecution and elimination of these cartels.³²

Not All Agreements Among Competitors Are Illegal

Although it is true that an agreement whose sole purpose is to fix prices or restrict output is a per se violation of the antitrust laws, it is not true that every agreement that results in prices being fixed is illegal per se. The Supreme Court has indicated that if the price fixing is ancillary to achieving another procompetitive purpose, then the agreement may well be deemed lawful. In such situations, it is necessary to investigate whether the price-fixing agreement is necessary to achieve the procompetitive purpose that underlies an agreement.

The Court long ago recognized that competitors sometimes must cooperate for the sake of efficiency and that this cooperation could involve pricing. Rather than applying a per se rule of illegality to such agreements, the courts apply a *rule of reason* analysis, in which the reasonableness of the cooperation is analyzed. One famous early case in which the rule of reason is eloquently articulated is *Chicago Board of Trade*.³³ In that

³¹*United States v. Socony-Vacuum Oil Co.*, 310 U.S. 150 (1940). This case is also referred to as *Madison Oil* because it was tried in Madison, Wisconsin.

³²In contrast to these findings, Sproul (1993) found that Department of Justice prosecutions between 1973 and 1985 had little effect on prices. However, Sproul used less detailed data than those used in the studies cited in the text.

³³*Board of Trade of City of Chicago v. United States*, 246 U.S. 231 (1918). *Standard Oil Co. of New Jersey v. United States* 221 U.S. 1 (1911) is the first case to employ a rule of reason.

case, members of the Board of Trade (who compete with each other to buy and sell contracts involving grains) agreed among themselves that after the Board had closed, no member of the Board of Trade could transact in a certain type of grain at a price other than the closing price that day. The Board of Trade was open during the early part of each day, and during that time, members transacted at prices that were determined by their willingness to buy or sell. The last price of the day was the closing price. The rule that no members could trade after the Board had closed except at the closing price made it more difficult for members to transact after closing, because supply and demand were likely to have moved the equilibrium price away from the closing price. The effect of this rule, therefore, was to create an incentive for members who wanted to trade to do so when the Board was open.

An organized exchange provides a valuable service. It amalgamates the information flows of buyers and sellers in such a way as to create a market price. An exchange is compensated for its activities by charging in some way for each trade that occurs. If one could costlessly observe the prices at the Board of Trade without having to pay any fees to it, one could free ride on the informational activities at the Board of Trade. By waiting to trade until after hours, one could use the information generated during the trading session by the Board of Trade yet avoid paying any fee. Therefore, this rule had two effects. First, it created an incentive to conduct more trades during the day on the exchange, making the market a larger one that can process more information. Second, it reduced the free-riding problem by discouraging trading after hours.

The Court ruled that this agreement was not a per se violation of the antitrust laws. The opinion, written by Justice Brandeis, said that

Every agreement concerning trade, every regulation of trade, restrains. . . . The true test of legality is whether the restraint imposed is such as merely regulates and perhaps thereby promotes competition or whether it is such as may suppress or even destroy competition.

Thus, the Court clearly believed that a cooperative agreement among rivals about pricing can promote competition sometimes.

In *Broadcast Music, Inc. (BMI)*, the Supreme Court investigated the way in which music is licensed.³⁴ Copyright owners of musical scores have property rights to their material. No one is allowed to use that material without permission and the payment of the agreed-upon fees. For example, any time a copyrighted song is played on the radio or on television, the copyright owner of that musical score must be compensated. It would be very costly for television and radio stations to locate and pay the copyright owner of each of the musical scores that it uses. Similarly, it would be very difficult for individual copyright owners to constantly monitor radio and television to determine if their musical scores were being performed.

To get around these horrendous transaction problems, two organizations were formed. One is the American Society of Composers, Authors, and Publishers (ASCAP) and the other is Broadcast Music, Inc. (BMI). Copyright owners belong to one

³⁴*Broadcast Music, Inc. v. Columbia Broadcasting System, Inc.*, 441 U.S. 1 (1979).

or both of these organizations and rely on them to collect revenues on their behalf. These organizations monitor musical productions and issue blanket licenses that enable the licensee to use any song listed in the blanket license. Fees for blanket licenses are ordinarily flat dollar amounts or percentages of total revenues. Therefore, ASCAP and BMI do fix prices in some sense, and they are, of course, organizations of competitors, handling the many songwriters.

The Supreme Court realized that ASCAP and BMI were providing an important service that lowered transaction costs and that the only way they could provide it was to set the price. In this sense, then, both BMI and ASCAP were performing procompetitive functions, and, by lowering transaction costs, they were expanding the amount of consumption that could occur. The Supreme Court therefore recognized that the per se rule was not appropriate here; instead, it decided that a rule of reason was necessary to analyze the reasonableness of the restraint. This case emphasized that cooperative agreements regarding price need not always violate the antitrust laws. (See Carlton and Klammer 1983, Halverson 1988, and Example 19.5 for an alleged price-fixing case involving not-for-profit colleges and universities.)

Information Exchanges Among Competitors

A common and natural form of association among competitors is a trade association, which is an organization composed of firms in similar businesses. Trade associations often collect information on the industry that is valuable to its members. Of course, trade associations can also serve as a vehicle by which prices are fixed (Chapter 5). However, it is important to recognize the legitimate information-generating services that trade associations can provide, such as revealing cost information to their members, or even revealing transaction prices to market participants, provided there is not collusion.

In the *Hardwood* case, the Court investigated the activities of the American Hardwood Manufacturers' Association, which had about 400 members.³⁵ The association engaged in gathering and reporting information about the sales, production, inventory, and pricing activities of each member and making such information available to the members. Moreover, at their meetings, members frequently discussed business conditions and the suitability of increases or decreases in industry production in light of these conditions. This behavior is consistent with that of a cartel. A cartel with 400 members, however, would be a difficult one to police and is therefore not likely to be successful in raising price for long. Therefore, it is likely that the exchanged information probably improved the knowledge of market conditions without increasing price. Nonetheless, the Court ruled that these activities were illegal (see Example 11.8).

Justice Brandeis disagreed with his colleagues on the Supreme Court about the *Hardwood* case. He explained that had there been a centralized market, much of the information collected by the trade association would have been automatically available. The provision of information was viewed by Brandeis as a beneficial, procompetitive effect of the trade association.

³⁵*American Column & Lumber Company v. United States*, 257 U.S. 377 (1921).

EXAMPLE 19.5*Colleges and Antitrust:
Does Your School Belong to a Cartel?*

In the 1950s, some Ivy League schools met and agreed not to offer aid to star athletes except on the basis of the financial need of the athlete. Soon, the agreement was extended to cover star students. By the 1980s, twenty-three elite schools in the Northeast were participating in this agreement, which was called the Overlap agreement. Under the Overlap agreement, each school would (1) agree to provide aid to students only on the basis of need, (2) adopt similar procedures to define need, and (3) meet to examine the actual awards made to each student who was admitted to at least two of the schools participating in the Overlap agreement and adjust aid offers if they differed.

The schools claimed that the Overlap agreement allowed them to conserve their financial resources to concentrate aid on poor students so as to achieve the twin goals of (1) having admission decisions based only on merit, and (2) guaranteeing full financial aid (based on need) to every admitted student. Few schools outside of those participating in the Overlap agreement adhered to these twin goals. The schools also claimed that their Overlap policies were fully consistent with federal education policies, which, for the most part, forbid the use of federal funds for scholarship aid if the aid is not based on need.

In 1991, the U.S. Department of Justice sued the eight members of the Ivy League and MIT under Section 1 of the Sherman Act. The Department of Justice claimed that the Overlap agreement was a per se illegal price-fixing agreement designed to raise each school's revenue. All the Ivy League schools agreed to stop the behavior and the suit against them was dropped. MIT refused to settle and went to trial.

Although colleges might want to cooperate to raise their own revenue and thereby harm students, they also might want to cooperate to achieve a social goal such as helping poor students. After all, one of the objectives of a not-for-profit college is to benefit students. Carlton presented an econometric study of average tuition paid by all students that revealed that there was no evidence to support the view that average tuition was higher as a result of the Overlap agreement. There was no question, however, that the Overlap agreement caused some students to pay higher tuition and others to pay lower tuition than they would have otherwise paid.

The court ruled that a per se approach should not be used because of the not-for-profit nature of colleges. The court found that the Overlap behavior did result in a violation of the antitrust laws because it restrained competition—specifically, the bidding for star students. Curiously, the Department of Justice did not attack the agreement not to bid for star athletes. The case was reversed on appeal and sent back to the district court for further review. Soon after the initial district court decision, Congress passed legislation to make it legal for the schools to continue to abide by most of the Overlap agreement that had been judged illegal. The case was settled with the schools being allowed to engage in most of the conduct covered by the Overlap agreement. However, the schools never reinstituted the Overlap conduct. Subsequent research by Hoxby (2000) confirmed the predictions of the original econometric model that Overlap did not cause an increase in average tuition.

Source: Carlton, Bamberger and Epstein (1995) and Bamberger and Carlton (2004). Carlton served as an expert witness for MIT.

A few years later, the Court again examined another trade association, the Maple Flooring Manufacturers' Association, with twenty-two members who accounted for roughly 70 percent of the total production of hardwood-type floors.³⁶ The association provided information on costs, freight, quantities sold, and prices received by individual members, and held meetings at which various industry members exchanged views about the state of the industry. The Court ruled that this activity was not a violation of the antitrust laws and cited the procompetitive benefits that result from a free flow of information and having industry participants apprised of market conditions. Using the economic theories about number of participants developed in our discussion of cartels, it appears that the trade association was much more likely to act successfully as a collusive device in the *Maple Flooring* case than it was in the *Hardwood* case. Despite this, the Maple Flooring Association was exonerated, but not the Hardwood Association.

Several decades later, the Court investigated exchanges of price information among producers of corrugated containers.³⁷ One competitor would request information from another on the most recent price that it had offered. The industry was concentrated, with the defendants accounting for about 90 percent of the shipments of corrugated containers from plants in the southeastern United States. After examining the economic factors of the industry including its concentrated structure, the Court concluded that the exchange of information was anticompetitive.

The discussion of oligopoly theory in Chapters 5 and 6 shows that exchanges of information can assist in collusion. For this reason, courts have paid careful attention to the activities of trade associations. At the same time, the courts recognize that information is a scarce commodity and that its dissemination can often be valuable. Evaluating these two offsetting effects is difficult.

Oligopoly Behavior

Noncooperative oligopoly prices may be above the competitive level because firms recognize their mutual interdependence and find it in their interests not to drive prices to competitive levels. The question the courts had to address was whether such pricing and other oligopoly behavior can be regarded as the result of an agreement among competitors that violates the antitrust laws. The enforcement of the antitrust laws often focuses on explicit agreements among competitors. The prosecution effort centers on showing evidence of an agreement (for example, incriminating documents) rather than on showing the effects of an agreement (for example, higher prices).

The Court addressed the question of when one could infer that a conspiracy or agreement had been made among competing firms in *Interstate Circuit*.³⁸ The Court said, "In order to establish agreement, it is compelled to rely on inferences drawn from the course of conduct of the alleged conspirators." The Court ruled that similarity in behavior was enough to constitute evidence of an agreement.

³⁶*Maple Flooring Manufacturers' Association v. United States*, 268 U.S. 563 (1925).

³⁷*United States v. Container Corp. of America*, 393 U.S. 333 (1969).

³⁸*Interstate Circuit, Inc. v. United States*, 306 U.S. 208 (1939).

In the *American Tobacco* case, the Court examined in detail the behavior of the cigarette industry in the 1930s.³⁹ List prices of the three major companies (the “Big Three”), Reynolds, American, and Liggett & Myers, were identical most of the time. During the height of the Depression, the cigarette companies all raised their prices even though their costs fell.

After prices rose, new competitors entered the cigarette industry and were able to sell their brands for 10¢, which was less than the 15¢ charged for the brands of the three majors. The market shares of the Big Three started to erode, and they had lost roughly 22 percent of total cigarette sales by 1932. The Big Three responded by cutting prices, and sales of the 10¢ brands fell considerably: The market share of the 10¢ brands was reduced to around 6.5 percent by 1933. The three major cigarette companies used their influence to make sure that no retail store sold the brands of the Big Three for more than 3¢ above the price of the 10¢ brands. See Example 11.2.

The Court found that the similarity of conduct among the three major companies provided a basis to infer that an unlawful conspiracy had occurred:

Where the circumstances are such as to warrant a jury in finding that the conspirators had a unity of purpose for a common design and understanding, or meeting of minds in an unlawful arrangement, the conclusion that a conspiracy is established is justified.

After *American Tobacco*, it was unclear exactly what type of oligopoly behavior would be subject to the antitrust laws. Was merely parallel behavior, in which firms who recognize each other’s interdependence act similarly, a violation of the antitrust laws?

In a series of cases, the Court aggressively attacked oligopoly behavior involving delivered pricing (see Chapter 11). However, in 1954 the Court indicated a change in direction. In the *Theatre Enterprises* case, the Court addressed the question of parallel behavior of movie theaters.⁴⁰ A newly refurbished theater sought to obtain the rights to run first-run feature movies from several distributors. The distributors refused because they already had theaters lined up for their first-run features. The Court ruled that

business behavior is admissible circumstantial evidence from which the fact finder may infer agreement . . . but this Court has never held that proof of parallel business behavior conclusively establishes agreement or, phrased differently, that such behavior itself constitutes a Sherman Act offense.

In other words, the common action of the distributors in refusing the movie theater the right to run first-run movies did not constitute a violation of the antitrust laws. This case is often interpreted to mean that parallel behavior (“conscious parallelism”), the kind that naturally results from a few firms’ competition with each other in an oligopoly, cannot by itself lead to an antitrust violation; there must be some additional offense (“conscious parallelism plus”) for the behavior to constitute an illegal action.

³⁹*American Tobacco Company v. United States*, 328 U.S. 781 (1946).

⁴⁰*Theatre Enterprises Inc. v. Paramount Film Distributing Corp.*, 346 U.S. 537 (1954).

The view that parallel behavior alone is not sufficient for an antitrust violation has been reaffirmed in several recent cases that the Federal Trade Commission has brought unsuccessfully, in which it has alleged that either markets are so-called shared monopolies (firms choose not to compete for the same customers and instead have local monopolies)⁴¹ or are not competitive because of certain business practices adopted independently by each firm. For example, in *du Pont*, the FTC charged that the non-collusive adoption of certain common business practices, such as notification to buyers of price increases, the use of a most-favored nations clause (see Chapters 5 and 11), the use of uniform delivered pricing, and public announcements in the press all constituted business practices that facilitated noncompetitive pricing.⁴² The Court of Appeals for the Second Circuit rejected such arguments as indicating violations of the antitrust laws:

The mere existence of an oligopolistic market structure in which a small group of manufacturers engage in consciously parallel pricing of an identical product does not violate the antitrust laws.

Mergers

Prohibitions against price fixing would have little effect without limits on mergers. The antitrust laws try to prevent the creation of additional market power through mergers of competitors. The issue in a merger case is not whether the industry is currently competitive, but whether it will become less competitive as a result of a merger. Because mergers can generate efficiencies, a merger policy that overdeters merger activity imposes a significant cost on society. Conversely, too lenient a policy leads to the creation of additional market power. We first discuss mergers among competitors and then among potential competitors.

Mergers of Competitors. In an early decision, *Northern Securities Company*, the Supreme Court investigated the creation of a holding company that would control two large, competing railroads: The Great Northern Railroad Company and the Northern Pacific Railway Company.⁴³ The creation of this holding company, which would exercise control over these two previously competing railroads, was deemed to violate the antitrust laws. The *Northern Securities* decision in 1904 coincided with the end of the widespread merger movement in the early 1900s (see Chapter 2).

Soon after the *Northern Securities* decision, the Court reached another decision involving market power acquired through merger. In *Standard Oil*, the Court investigated the creation of the Standard Oil Company and the practices it followed in

⁴¹*FTC v. Kellogg et al.*, Docket No. 8883, 99 FTC Reporter 8, 1982. The FTC eventually dismissed the case. See Schmalensee (1978b) for an analysis of this case.

⁴²*E.I. du Pont de Nemours & Co. v. FTC*, 729 F.2d 128 (2d Cir. 1984). This case is also sometimes called the *Ethyl* case because Ethyl was a participant. See Example 11.7 and Hay (1999).

⁴³*Northern Securities Company v. United States*, 193 U.S. 197 (1904).

acquiring businesses related to petroleum products.⁴⁴ John D. Rockefeller and others were the defendants. One charge was that the defendants

purchased and obtained interest . . . and entered into agreements with . . . various persons . . . engaged in purchasing, shipping, refining, and selling petroleum and its products . . . for the purpose of fixing the price of crude and refined oil and the produce thereof, limiting production thereof, and controlling the transportation therein, and thereby restraining trade . . . and monopolizing interstate commerce.

Another charge was that refineries that refused to enter into the agreement were driven out of business through a variety of predatory tactics such as low prices. Other charges included unfair practices against competing pipelines, contracts with competitors, espionage, and division of the United States into districts and limiting the amount of competition in each district. The Court ruled that the actions indicated “a conviction of a purpose and intent” to monopolize, and it ordered the dissolution of the combination. This case is famous because the Court refused to apply a per se ban to mergers among competitors and introduced the rule of reason, in which one had to investigate whether the resulting effect of the merger was an unreasonable restraint of trade.

In the *United States Steel* case, the Court seriously retreated from vigorously applying the antitrust laws to enjoin merger activity.⁴⁵ The case involved the creation of the United States Steel Company through the merger of approximately 180 independent firms. U.S. Steel produced 80 to 90 percent of the entire steel output of the country. The Court refused to find the creation of U.S. Steel illegal, and seemed to indicate that because U.S. Steel, unlike Standard Oil, did not engage in improper behavior, the combination was lawful.

Dissatisfaction with the Supreme Court’s treatment of mergers (especially in light of a failure to block another acquisition)⁴⁶ led Congress to pass the *Celler-Kefauver Act* in 1950, which strengthened Section 7 (on merger activity) of the Clayton Act. In *Brown Shoe*, the Supreme Court applied the new standards of the amended Section 7 of the Clayton Act to block a proposed merger between G.R. Kinney Company and Brown Shoe Company.⁴⁷ Both were manufacturers and retail sellers of shoes. The language of the Court’s decision indicated that a combined share of 5 percent in a city was excessive, taking into account the trend toward increasing concentration in this industry. The Court also issued its famous laundry list of criteria for defining a market, which we discussed in the section on market definition.

The Court continued its hard line on mergers by stopping a merger among banks in *Philadelphia Bank*.⁴⁸ The merged firm would have had less than 40 percent of de-

⁴⁴*Standard Oil Company of New Jersey v. United States*, 221 U.S. 1 (1911). McGee (1958) analyzes this case.

⁴⁵*United States v. United States Steel Corporation*, 251 U.S. 417 (1920).

⁴⁶*United States v. Columbia Steel Company*, 334 U.S. 495 (1948).

⁴⁷*Brown Shoe Company v. United States*, 370 U.S. 294 (1962). Peterman (1975) analyzes *Brown Shoe*.

⁴⁸*United States v. Philadelphia National Bank*, 374 U.S. 321 (1963).

posits in the Philadelphia area. The Court also rejected a consideration of the efficiency benefits of a merger.

The Supreme Court took its strictest stance in enforcing Section 7 of the Clayton Act in *Von's*.⁴⁹ Von's Grocery Company sought to acquire Shopping Bag Food Stores, another retail grocery company operating in Los Angeles. Their combined sales accounted for only 7.5 percent of all sales in Los Angeles, yet the Supreme Court prevented this acquisition. Shortly thereafter in 1968, the Department of Justice issued very strict guidelines on which firms could likely merge without challenge (see www.aw-bc.com/carlton_perloff "1968 Merger Guidelines").

The more recent 1984, 1992, and 1997 Merger Guidelines (Example 19.3) recognized the potential efficiency gains from mergers. These guidelines apparently are a response to the earlier rejection by the government and the Court to using proposed efficiency gains from mergers as a defense to justify a merger that increases concentration in a market. The application by the Department of Justice and the FTC of the current merger guidelines, which recognize the value of efficiencies, suggests that efficiencies alone generally do not provide sufficient justification for a merger in which prices are expected to rise. Efficiencies, however, can provide a justification for a merger that results in increased concentration if the efficiencies would lead to lower prices.

One defense that courts have allowed in merger cases is the *failing-firm defense*, in which the firms explain that if the proposed merger is not allowed, one of the firms will go out of business. If the proposed transaction is the least anticompetitive one that can prevent the assets from leaving the industry, the Department of Justice or FTC will not challenge the merger. However, if the failing firm goes bankrupt but the creditors continue to operate the firm, then the bankruptcy does not affect competition and there is no reason to allow a failing-firm defense.

The failing-firm defense can be regarded as a recognition that current market shares may not reflect the future importance of the competitor that will vanish as a result of the merger. If a firm will go out of business unless it merges with others, then the fact that it currently has a high market share is irrelevant in considering whether the merger should go through or not. Merger policy should be forward-looking, and it is really the future competitive significance of the merging firms that is important in understanding whether a merger is anticompetitive.⁵⁰ This principle is recognized by the Supreme Court in several cases in which it finds that current market shares may be inaccurate indicators of the future competitive significance of a firm.⁵¹

⁴⁹*United States v. Von's Grocery Company*, 384 U.S. 270 (1966).

⁵⁰Suppose that there are three firms in an industry with market shares of 30%, 30%, and 40%, and that the one with 40% is failing. If it fails, the remaining firms will have 50% and 50%. If, instead, the failing firm is acquired by one of the remaining firms, the shares will be 70% and 30%. Thus, if a merger occurs, concentration as measured by the Herfindahl-Hirschman Index (HHI) increases (see Chapter 8). The comparison of the HHIs is *irrelevant* however, if the output level is not held constant. If, as a result of the acquisition, more assets remain in the industry and output is permanently higher, then consumers are better off even if market concentration increases from what it would have been if no acquisition had occurred, the failing firm had failed, and its assets had exited the industry.

⁵¹*United States v. General Dynamics Corporation*, 415 U.S. 486 (1974).

Thus, the criteria for analyzing mergers used by the Court and the government have evolved considerably since the Celler-Kefauver Act in 1950. As the Court has eliminated some of the inconsistencies in its opinions in defining markets, and as economists and lawyers have become more sophisticated about defining markets and understanding the effects of market concentration, government policy toward mergers has become more systematic. A merger such as that attacked in *Von's* would probably not be attacked today. Moreover, both the FTC and Department of Justice address their concerns about lack of competition resulting from a proposed merger by allowing the merging firms to restructure the proposed transaction to remedy competitive concerns (such as by selling some assets to a new entrant). They have used this “fix it first” policy extensively since the 1980s.

Mergers of Potential Competitors. Suppose two firms that do not currently compete in the same market wish to merge. Can the merger be blocked if the government thinks it is likely that the two firms would have competed in the future? Logically, there is nothing wrong with blocking a merger if it will improve future competition. Practically, it is very difficult to determine which firms are potential competitors. The decisions of the Court have evolved over time so that a merger between potential competitors is now much less likely to be challenged as anticompetitive.

An early case involving a merger between potential competitors was *El Paso Natural Gas*.⁵² El Paso Natural Gas sought to acquire the assets of Pacific Northwest Pipeline Corporation. Both companies operated large natural gas pipelines. Only one of them, El Paso, delivered natural gas into California, a market in which the government contended competition would be lessened if the acquisition occurred. Even though Pacific Northwest had never sold gas in California, on several occasions it had attempted to obtain the necessary regulatory approval to deliver gas into California. Indeed, Pacific Northwest had conducted lengthy negotiations with a large customer in southern California. The result of these negotiations was to heighten competition, even though El Paso eventually won away the customer. The Court ruled that although Pacific Northwest was not a successful seller in California, it was indeed a competitor: “Unsuccessful bidders are no less competitive than the successful ones.” The acquisition was barred. Because Pacific Northwest had actually bid for business, it seems more reasonable to regard this case as one between actual rather than potential competitors.

Another important case involving potential competition was *Procter & Gamble*.⁵³ Procter & Gamble Company acquired Clorox Chemical Company, which was the leading manufacturer of household liquid bleach and had about 50 percent of U.S. sales. Procter & Gamble did not manufacture or sell bleach, but was a major manufacturer and seller of many other household products. The Court decided that the acquisition should be blocked because Procter & Gamble was a likely entrant into the liquid-bleach market. As a result of the decision, Clorox was divested in 1969 (12 years after the original merger).

⁵²*United States v. El Paso Natural Gas Company*, 376 U.S. 651 (1964).

⁵³*Federal Trade Commission v. Procter & Gamble Company*, 386 U.S. 568 (1967).

The potential competition doctrine was again used by the Supreme Court in *United States v. Falstaff Brewing Corporation*.⁵⁴ Falstaff, one of the nation's largest brewers, sought to acquire Narragansett, which was the largest brewer in the New England area—an area in which Falstaff did not compete. The government argued that the merger should be enjoined because Falstaff was a likely entrant into the New England area. The district court had found that Falstaff had no intentions of otherwise entering the New England area, but the Supreme Court ruled that Falstaff may have affected competition in New England anyway because Falstaff might have been perceived as a potential entrant into the New England area and therefore might have restrained prices. The district court subsequently found that Falstaff was not perceived as a potential entrant. Although logically a perceived potential competitor could influence the market, a perceived potential competition doctrine depends on the state of mind of the competitors and not on any easily verifiable facts. Therefore, even though the doctrine is logically consistent, it turns out to be extremely complex to litigate such a case, which is based on the opinions of competitors who may be interested in preventing a merger that would result in the creation of an efficient competitor and rival.

The Supreme Court significantly constrained the application of the potential competition theory in *Marine Bancorporation*.⁵⁵ The government challenged a proposed merger between a commercial bank in Seattle and one in Spokane: two banks that were not direct competitors even though both were located in the same state. The government challenged the merger on the grounds that the acquiring bank would have found an alternative and more competitive means for entering the Spokane area. The Court was unconvinced that an alternative method of entry would achieve the same procompetitive effects as this acquisition. It appears, then, that the Court's decision requires a showing that first, the potential competitor has some unique advantage to entry, and second, that this means of entry would allow the potential competitor to enter and prosper. Since the *Marine Bancorporation* decision, the potential competition doctrine has not fared well (Posner and Easterbrook 1980, 531). Of course, if markets are broadly defined, there is little need for a potential competition doctrine because the potential competitors are considered part of the market.

Exclusionary Actions and Other Strategic Behavior

So far, this chapter has described how antitrust laws are designed to prevent agreements between competing firms, such as a price-fixing conspiracy or a merger, that can lead to the creation of market power. This section examines actions by a single firm (or firms acting collectively) that may help it maintain its monopoly or facilitate its acquisition of market power at the expense of its rivals. These **exclusionary actions** are used

⁵⁴*United States v. Falstaff Brewing Corporation*, 410 U.S. 526 (1973). See also *United States v. Penn-Olin Chemical Co.*, 378 U.S. 158 (1964).

⁵⁵*United States v. Marine Bancorporation Inc.*, 418 U.S. 602 (1974).

by a firm to eliminate rivals from a market or harm them, thereby either helping to maintain or create a monopoly. These actions, or bad acts, include predatory pricing, denial of key products to rivals, vertical relationships among firms, and tie-in sales. Many of these practices are a violation of Section 2 of the Sherman Act. Hence, antitrust cases alleging these actions are often called *Section 2 cases*.

Section 2 of the Sherman Act forbids firms' exclusionary conduct (bad acts) that adversely affect competition. A recurrent problem in Section 2 cases is that the Court has been unclear exactly how vigorously a dominant firm can respond to new competition. Moreover, economists cannot usually say with certainty which types of strategic behavior lead to benefits for consumers when competitors are harmed (Chapter 11). For example, a firm may strategically invest before other firms can enter an industry. Such a policy can benefit consumers even if it prevents potential competitors from entering the market. Thus, blanket prohibitions of such behavior may be harmful in some industries.

Section 2 litigation can be costly (as can all complicated litigation). One example of costly litigation is the IBM case in which the government sought to force IBM to break itself up into several firms. The government claimed that IBM practiced numerous policies designed to exclude competition. The legal fees as well as the time of IBM and government employees probably put the litigation cost in the area of hundreds of millions of dollars. The government eventually dropped the case.⁵⁶

How vigorously should courts use Section 2 to constrain the action of firms? The answer turns on specific attributes of a market. Where entry can occur quickly, market power may be short lived, and there may be no need for Section 2 litigation. Over-vigorous enforcement of Section 2 cases, in addition to reducing market power, could dissuade firms from pursuing certain efficient policies that would benefit consumers. This efficiency loss could be large and would not diminish over time. Striking the right balance in Section 2 cases remains a difficult problem for the courts. We now examine strategic behavior by a firm with respect to both rival firms in its market and vertical relations.

Competition Between Rivals

In general, competition benefits consumers; however, some forms of competitive behavior can reduce competition, as Chapter 11 shows. This section reviews some of the main types of behavior between rivals that the Court has found to violate the antitrust laws. We begin with a general discussion of some famous cases in which the Court deemed certain behavior undesirable and then discuss the specific examples of predatory pricing and denial of key products to rivals.

Competitive Behavior Deemed Undesirable by the Court

One of the most famous Section 2 cases is *Alcoa*.⁵⁷ Alcoa produced and sold aluminum ingot and also fabricated the aluminum ingot into many finished and semifinished goods. In part because Alcoa owned or licensed many of the critical original

⁵⁶See Fisher, McGowan, and Greenwood (1983) for a description of this lengthy litigation from IBM's perspective. See Houthakker (1985) for a different view.

⁵⁷*United States v. Aluminum Company of America*, 148 F.2d 416 (1945).

patents, no firm could effectively compete with Alcoa prior to 1909. In the government's 1945 case, it alleged that, after 1909, Alcoa maintained its market power through a series of exclusionary tactics, among them (1) the signing of power contracts that forbade the power companies to sell power to anyone else who made aluminum; (2) explicit price-fixing agreements with foreign producers of aluminum to prevent imports into the United States; (3) a price squeeze, in which the price of aluminum ingot was raised to independent aluminum-sheet fabricators, who were then unable to make a profit fabricating the sheet and selling it in competition with Alcoa at the prices Alcoa was setting for aluminum sheet; and (4) a strategy of expanding capacity with the intention of eliminating competition. Alcoa remained the sole domestic producer of aluminum until 1945.

One of Alcoa's defenses was that the profit it earned was not very high. The court ruled that whether profits are high or low is irrelevant: "[Congress] did not condone good trusts and condemn bad ones; it forbade all."⁵⁸ The court stated that *the mere acquisition of a monopoly by itself was not necessarily illegal*.

Despite this view, which implies that efficient firms that grow should not be penalized, the court looked with disfavor on Alcoa's policy of anticipating demand and building capacity for it in advance:

It was not inevitable that it should always anticipate increases in demand for ingot and be prepared to supply them. . . . It insists that it never excluded competitors; but we can think of no more effective exclusion than progressively to embrace each new opportunity as it opened.

The court's reasoning is perplexing. It is difficult for an economist to distinguish evil capacity expansion from desirable capacity expansion that occurs as a result of foresight.

The court also ruled that "The monopolist must have both the power to monopolize, and the intent to monopolize." By stressing intent, the frame of mind of the violator becomes relevant in an antitrust suit. Endless litigation can result when someone's frame of mind, rather than the actual effects of the economic actions, is the subject of the litigation. The court further ruled that Alcoa's price-squeeze policy was unlawful.⁵⁹

⁵⁸The *Alcoa* case was decided by a court of appeals rather than the Supreme Court; the Supreme Court was unable to hear the case because of a conflict of interests involving several of the Justices. The Court of Appeals for the Second Circuit was designated as the court of last resort for the *Alcoa* case, and Judge Learned Hand wrote the decision.

⁵⁹A fascinating issue in the *Alcoa* case was the definition of the market. Aluminum ingot, once it is made into fabricated aluminum, can be recycled as scrap aluminum. Scrap aluminum competes with primary ingot for many uses. The question arose as to whether the secondary market should properly be considered as part of the market in which virgin aluminum ingot competes. The court ruled that secondary aluminum should not be part of the market definition and concluded that a market share for virgin ingot of 90 percent would definitely indicate monopoly power; 67 percent might indicate monopoly power; but 33 percent would not. Secondary and primary products definitely compete with each other, but such competition need not erode the initial market power in the primary product. Once the primary product is sold, there may be no further monopoly profits to be made, because the secondary market does constrain the *subsequent* pricing of primary aluminum, even though it does not constrain the initial price. (See Chapter 15, "Renting Versus Selling by a Monopoly.") Another issue in the definition of the market involved whether imports should be included in the market. The court correctly decided to include them.

The Court remanded the case to the district court for reconsideration. The main antitrust divestiture order facing Alcoa, resulting from the antitrust litigation, was one regarding its Canadian properties. During the time period of the district court's reconsideration, the United States government sold off aluminum facilities built for it during World War II and thereby set up Reynolds and Kaiser as two competitors to Alcoa. The monopoly on aluminum that Alcoa enjoyed in the United States disappeared. By 1958, Alcoa's share of primary aluminum ingot capacity had fallen to 35 percent.

The *Alcoa* decision had far-reaching implications for dominant firm behavior. It was unclear if there was anything a dominant firm could do to avoid being charged with consciously seeking to maintain control of the market. Further, it was unclear how one would determine whether its monopoly condition was "thrust upon it," maintained by clever but legal business practices, or maintained by practices that the courts would find illegal.

In *United Shoe*, another major Section 2 case, the government charged that United Shoe maintained its market share of 75 to 85 percent of American shoe machinery primarily through the practice of refusing to sell its equipment, agreeing only to lease it.⁶⁰ The government maintained that United Shoe, by only leasing its equipment, created barriers to entry. The reason was that because United repaired its own equipment, there were no independent repair organizations that a competitor could rely on; therefore, if a competitor sought to enter the field, it would have to also provide repair services. The Court also ruled that the leasing system under which United leased the machines for 10 years would "deter a shoe manufacturer from disposing of a United machine and acquiring a competitor's machine." The Court ruled that the leases were "so drawn and so applied as to strengthen United's power to exclude competitors." Although the Court recognized the superiority of many of United's products and services, it felt that the leasing system contributed to its market power. The Court required United to offer for sale any machines that it leased. *United Shoe* appears to illustrate the important concept presented in Chapter 15 that a monopoly would prefer to lease rather than sell its machines; however, see Example 15.1.

The Court's views on the 10-year period of the lease are troublesome. If leases come up for renewal over time, and if there can be competition to obtain the customer whose lease has expired, then it is unclear why competition is reduced by the leases. Only if the slow turnover of customers prevents a rival from attaining some critical mass necessary for its survival as an efficient competitor would there seem to be an antitrust concern.⁶¹ Even in that case, one would also want to consider any benefits that arise as a result of the long-term nature of the contract.

In *Griffith*, the Court considered the buying practices of chains of motion picture theaters.⁶² These motion picture theaters paid for the movies through rentals that were based on the total attendance of the entire chain, rather than at any particular

⁶⁰*United States v. United Shoe Machinery Corporation*, 110 F. Supp. 295 (1953). This citation refers to the Federal Supplement, a standard legal reference in which the opinions of the district courts appear.

⁶¹Aghion and Bolton (1987) analyze models in which long-term contracts can create anticompetitive harm by allowing the buyers to act collectively as a monopsony.

⁶²*United States v. Griffith*, 334 U.S. 100 (1948).

theater. That meant that if a chain had a theater in a town in which it was competing with a single, independent theater, the chain could obtain the same movie at a lower price than the single theater. The Court ruled that this placing of single competitors at a disadvantage was a violation of Section 2 of the Sherman Act. The Court also ruled that the effect of the action rather than the intent of the actor was a reasonable focus of inquiry.

In *Berkey*, the Court of Appeals for the Second Circuit examined the duty that a monopoly has toward its rivals.⁶³ In 1972, Kodak, the dominant firm in the markets for cameras and for film, introduced the 110 pocket Instamatic camera and a film format to fit that camera. Berkey was a manufacturer of cameras and a processor of film. One of Berkey's claims was that because Kodak refused to predisclose the format of its 110 film, Berkey was unable to manufacture cameras to fit the 110 format film until well after its introduction. Berkey claimed that Kodak's dominance in both film and cameras required it to predisclose to its competitors any changes in film format that would affect competition in the camera market. The court ruled that predisclosure was not a duty imposed on a dominant firm by the antitrust laws. The court recognized that the antitrust laws, especially Section 2, do not forbid monopolies. The court reiterated that the standard for a Section 2 offense is the possession of market power and the willful acquisition or maintenance of that power, as distinguished from growth or development as a consequence of a superior product, business acquisition, or historic accident.⁶⁴

Predation. One of the classic bad acts is predatory pricing (Chapter 11); however, there is a danger of confusing predatory pricing with aggressive competition. The *Utah Pie* case involved a claim of predatory pricing.⁶⁵ Utah Pie Company sold frozen dessert pies in Utah. Continental Baking Company, Carnation Company, and Pet Milk Company sold pies in competition with Utah Pie. The Salt Lake City market was the scene of dramatic price competition, and there was evidence to show that prices of the defendants' products were lower in Salt Lake City than they were elsewhere. Evidence suggested that the prices of the defendants' products, at least some of them, were below their direct cost plus an allocation for overhead. There was evidence that one of the defendants had employed an industrial spy to infiltrate the Utah Pie plant to obtain information. The Court ruled that such price discrimination eroded competition and, therefore, was predatory and in violation of the law.⁶⁶

In *Telex*, the Court of Appeals for the Tenth Circuit investigated IBM's pricing behavior with regard to peripheral devices (such as disk drives) that plugged into an IBM

⁶³*Berkey Photo, Inc. v. Eastman Kodak Company*, 603 F.2d 263 (2d Cir. 1979) cert. denied, 444 U.S. 1093 (1980).

⁶⁴See also *United States v. Grinnell Corp.*, 384 U.S. 563 (1966).

⁶⁵*Utah Pie Company v. Continental Baking Company*, 386 U.S. 685 (1967). Elzinga and Hogarty (1978) provide an economic analysis of this case.

⁶⁶The language of *Utah Pie* suggests that price discrimination can violate Section 2, even if prices exceed average cost. As discussed in Chapter 11, fully allocated cost is an inappropriate standard to use in determining if prices are predatory.

central processing unit.⁶⁷ Telex claimed that IBM violated the antitrust laws by its decision to slash prices on its peripheral devices in order to compete with Telex. The court found that because the price was not below IBM's production costs, there were no grounds to the complaint.

In *Matsushita*, the Supreme Court again investigated a charge of predatory pricing.⁶⁸ This case involved a claim that certain Japanese manufacturers engaged in predatory pricing over a 20-year period. The Court recognized the irrationality of such a scheme—it would obviously be unprofitable to lose money for 20 years—and dismissed the case (Example 11.1). Alleged predatory behavior must be credible to be found to violate the law.

In *Brooke Group v. Brown and Williamson Tobacco* (113 S. Ct. 2578 (1993)), Liggett (Brooke), which pioneered the development of low-price generic cigarettes, charged that Brown and Williamson introduced its generic cigarettes at predatory prices. The Court held that a successful predation claim required proof that price was set below some measure of cost and the alleged predator had a reasonable likelihood of recouping its losses from predating. The Court found that the market structure for the sale of generic cigarettes would not allow Brown to recoup any predatory losses. The Court ruled that with no possibility of recouping its losses, even below-cost pricing does not support a claim of predatory pricing. Because the parties agreed to use average variable cost as the measure of costs, the Court declined to rule on what the appropriate cost measure should be in predation cases. Thus, there still is no Supreme Court precedent on that important issue.

Refusals to Deal and Essential Facilities. When a group of firms collectively decide to boycott or refuse to deal with a rival, thereby denying the rival access to certain markets, their actions can violate Section 1 in addition to Section 2. For example, in *Eastern States*, the Court condemned the actions of a group of retail lumber dealers who refused to deal with any wholesale lumber dealers who also sold at retail.⁶⁹

Frequently, the courts have treated collective action involving a refusal to deal with certain firms as a per se violation. Two recent cases may indicate a change in view. In *Northwest Wholesale Stationers, Inc. v. Pacific Stationery and Printing Company*, the Court refused to apply the per se rule to a case involving an agreement among competitors.⁷⁰ This case involved the expulsion of one member from a cooperative buying agency (a group of firms that buy products as one purchaser). The Court determined that the cooperative buying agency, through an agreement among competitors, did not necessarily engage in a per se violation by expelling the plaintiff and refusing to deal with that firm. The Court ruled that in the absence of proof that the cooperative

⁶⁷*Telex Corp. v. International Business Machines Corp.*, 510 F.2d 894 (1975).

⁶⁸*Matsushita Electric Industrial Co. v. Zenith Radio Corporation*, 106 S. Ct. 1348 (1986).

⁶⁹*Eastern States Retail Lumber Dealers Association v. United States*, 234 U.S. 600 (1914).

⁷⁰*Northwest Wholesale Stationers, Inc. v. Pacific Stationery and Printing Company*, 105 S. Ct. 2613 (1985).

had market power or unique access to a critical resource necessary for effective competition, it was not appropriate to treat the conduct as a per se violation, and instead it had to be subjected to a rule of reason.⁷¹

In deciding refusal to deal cases, courts often emphasize the role of **essential facilities**: scarce resources that a rival needs to use to survive. For example, a trucking firm that owns the sole bridge leading to an island owns a facility that is essential to rival trucking firms that deliver to the island. Under the essential facilities doctrine, the owner of the essential facility must sometimes make the facility available to competitors.

In *Terminal Railroad*, all the railroad bridges in St. Louis were owned by a group of railroads.⁷² The concern was that this control could allow the owning railroad companies to harm rival railroads (Reiffen and Kleit 1990). The Court ruled that the owning group had to provide access to rival railroads on reasonable terms.

Collective action receives close scrutiny under our antitrust laws because of the danger that competitors will agree to restrict competition. For that reason, many joint ventures that limit entry (such as sports leagues) face antitrust concerns regarding exclusionary acts that would not arise in the context of a single firm. Under *Colgate*, a single firm supposedly can decide with whom it deals.⁷³ But that doctrine has not always been followed and even single firms may have a duty to deal with rivals. For example, in *Aspen Ski Company v. Aspen Highland Skiing Corporation*, 472 U.S. 585 (1985), the Court ruled that an owner of three ski mountains had to continue its historical practice of cooperating with the owner of a fourth mountain in issuing lift tickets that allow skiers access to all four mountains. The reasoning seems to put a higher burden on a monopoly that once dealt with a rival than one that never did. In *Kodak*, the Court reiterated its view in *Aspen* that a monopoly may refuse to deal with its rivals “only if there are legitimate competitive reasons for the refusal.”⁷⁴ See Carlton (2001) for further analysis.

The context of many cases involving refusals to deal and essential facilities is that one firm owns a scarce resource that its rivals need in order to compete. These cases, therefore, have a vertical element and can best be viewed in the context of the models of Chapter 11 involving raising rivals’ costs or the natural advantage of an incumbent. For example, by denying access to or by raising the toll on the only bridge to an island, the railroad that owns the bridge can put its rival at a competitive disadvantage.

There are two noteworthy features about forcing one firm to provide supplies to its rival. First, the Court must be concerned that the firm with the scarce resource does not charge too high a price; otherwise, no rival will be able to compete even if it has

⁷¹See also *Federal Trade Commission v. Indiana Federation of Dentists*, 106 S. Ct. 2009 (1986).

⁷²*United States v. Terminal Railroad Association of St. Louis*, 224 U.S. 383 (1912). See also *Otter Tail Power Co. v. United States*, 410 U.S. 366 (1973).

⁷³*United States v. Colgate & Co.*, 250 U.S. 300 (1919).

⁷⁴*Eastman Kodak Co. v. Image Technical Services, Inc.*, 112 S. Ct. 2091 (1992) n. 32. Recent cases such as the FTC’s suit against Intel and the Department of Justice’s case against Microsoft have raised antitrust concerns about the way in which an allegedly powerful firm can deal with its customers when those customers are also its rivals in some products.

access to the scarce resource. Second, a firm with monopoly power is usually allowed to charge any price it likes. It is unclear why that principle should be different here for the single firm that owns the scarce resource just because the scarce resource is an input for its rival.

Vertical Arrangements Between Firms

So far, the section has described how one firm (or a group of firms acting collectively) can harm a competitor through bad acts. The antitrust laws also characterize certain types of vertical relationships among noncompeting firms, typically a manufacturer and a distributor, as bad acts that harm competitors. We now analyze vertical integration and vertical restraints (resale price maintenance, exclusive territories, and exclusive dealing). The area of the law dealing with vertical relationships has changed significantly, and certain aspects of the law appear to be inconsistent with economic theories. See Carlton (2001) for a detailed analysis.

Vertical integration and vertical restrictions are not necessarily anticompetitive (Chapter 12). Even when a manufacturer is a monopoly, it is not at all clear that the vertical restrictions it may impose on a distributor reduce consumer welfare. It is not possible to prove definitively that vertical integration or restrictions always improve society's welfare, but neither is it possible to prove that a monopoly's choice of quality or any other product dimension always improves consumer welfare. Moreover, it is typically costly to examine a particular case of a vertical relation or quality choice, and, even after lengthy examination, it may still be difficult to reliably predict the effect of the vertical relation or quality choice on consumer welfare. Few argue that the antitrust laws should be used to control how a monopoly chooses quality or manufactures its product, yet there is usually no greater justification for interfering in the monopoly's choice of distribution than for interfering in its choice of quality or production.

In some markets, however, vertical integration or restrictions reduce competition and harm society, and those are the ones the antitrust laws should try to prevent. We now discuss those situations.

Exclusive dealing can harm society if it prevents or impedes rivals from obtaining distribution of their product. The same is true for vertical integration into distribution. However, as long as other efficient methods of distribution are available to rivals, neither exclusive dealing nor vertical integration restrain the entry of rivals.

Other vertical restrictions, such as exclusive territories, can have anticompetitive effects if they are forced on a manufacturer by a dealer cartel. That is, the exclusive territories could be part of an agreement among competing dealers on how to allocate territories.⁷⁵ Only if the dealers have monopsony power, however, does such a claim make sense (Chapter 12). No manufacturer would willingly take part in such a conspiracy of its dealers, because it would raise the manufacturer's distribution costs.

Antitrust policy toward vertical relationships has implications for the ability of a firm to price discriminate. Vertical integration and restrictions may enable price dis-

⁷⁵Vertical integration and vertical restrictions can also be used by manufacturers to facilitate collusion by making it easier to detect cheating on a cartel of manufacturers (see Chapter 5).

crimination to take place (Chapter 9). For example, a manufacturer that wants to charge different prices in New York and California could do so if it prevents resale between the two states by requiring its distributors to sell only in their own territories. Because it is unclear whether society is, in general, harmed or helped by imperfect price discrimination (Chapter 9), it seems unwise to apply a per se ban. Moreover, to examine every instance of price discrimination under a rule of reason would be costly, and even after the analysis, it probably would be difficult to predict reliably the welfare effects of the discrimination.

Vertical Integration: The Court's early views on vertical relations in general and vertical integration in particular were unclear at best. Apparently, the Court was concerned with *foreclosure of competition*. For example, if a firm that manufactures shirts vertically integrated backward into producing buttons, the firm would have foreclosed competition in the button market because other button manufacturers could now no longer sell to that firm.

In *Yellow Cab*, the Court suggested that vertical integration through merger might be per se illegal.⁷⁶ However, soon thereafter, it reached the opposite (and more reasonable) conclusion in *Columbia Steel* that "it is clear to us that vertical integration, as such without more, cannot be held violative of the Sherman Act."⁷⁷

The next major vertical integration case was *du Pont*.⁷⁸ Since 1920 (or earlier), du Pont, a major supplier to General Motors of automotive finishes and fabrics, had owned a 23 percent stock interest in General Motors. The U.S. government brought suit, claiming that the vertical relationship violated the antitrust laws. Although it was unclear how consumers would be adversely affected by this vertical ownership, the Court ruled that du Pont's ownership violated the antitrust laws.

It appears that since the *du Pont* case, enforcement policy toward vertical mergers has been in line with the reasonable economic logic of *Columbia Steel*. The vertical guidelines of the Department of Justice (repudiated by the Clinton administration) emphasized that vertical integration alone is not objectionable; instead, they focus on whether the vertical integration could be used to increase market power.

Although the FTC and the Department of Justice are less hostile to vertical integration than in the past, the courts have not always concurred. For example, in *Fotomat*, the Court of Appeals for the Seventh Circuit ruled against a franchisor that sought to open outlets in competition with its own franchisee.⁷⁹ It is difficult to understand why vertically integrating forward into distribution is an antitrust violation simply because independent dealers face additional competition.

Vertical Restraints: Using contracts, a firm may impose vertical restraints on another firm instead of vertically integrating to directly control that firm. Were vertical

⁷⁶*United States v. Yellow Cab Co.*, 332 U.S. 218 (1947).

⁷⁷*United States v. Columbia Steel Company*, 334 U.S. 495 (1948).

⁷⁸*United States v. E.I. du Pont de Nemours & Company*, 353 U.S. 586 (1957).

⁷⁹*Photovest v. Fotomat Corp.*, 606 F.2d 704 (7th Cir. 1979).

restraints to be outlawed but vertical integration allowed, firms would have an increased incentive to vertically integrate. Important vertical restraints include resale price maintenance, exclusive territories, and exclusive dealing.

Resale Price Maintenance: A manufacturer may set a minimum (or maximum) price that retailers may charge, called *resale price maintenance*, because the manufacturer wants to control the retail price at which its product is sold to consumers (Chapter 12). In 1911, the Court addressed whether a manufacturer could place pricing restrictions on its distributors. In *Dr. Miles*, John D. Park, a distributor, refused to enter into a contract that established minimum prices at which Dr. Miles's drug products could be sold.⁸⁰ The Court ruled that this pricing agreement was illegal because it suppressed competition among dealers and was equivalent to the fixing of price.

This ruling was unpopular, and the antitrust laws were eventually amended to allow resale price maintenance for certain products. In 1937, Congress passed the Miller-Tydings Resale Price Maintenance Act and in 1951, the McGuire Act. These acts gave manufacturers the right to set retail prices free of any antitrust liability provided the states had a *fair-trade* statute that allowed resale price maintenance (which would allow products to be sold at a "fair" price). Many states passed such fair-trade laws. In states without fair-trade laws, it was easier for one distributor to free ride on the promotional efforts of other distributors, because resale price maintenance is one way to control free riding (Chapter 12). The laws allowing resale price maintenance were repealed in 1975, and all resale price maintenance again became per se illegal.

The procompetitive logic of resale price maintenance is that resale price maintenance is one way for a manufacturer to induce its distributors to promote its products (Chapter 12). This logic was not understood by most economists prior to the 1960s. Since the 1960s, economists have discussed the competitive benefits of restraints that manufacturers want to place on the distributors of their products. These economists make no distinction between pricing restrictions and other restrictions that manufacturers might want to place on their distributors. Both can promote competition and prevent free riding (Posner 1981). Resale price maintenance can be anticompetitive, however, if it facilitates collusive behavior.

With the repeal of the laws permitting resale price maintenance in 1975, manufacturers can no longer set price floors for distributors.⁸¹ However, a recent decision by the Supreme Court may indicate relaxation of this ban.⁸² The Court analyzed a case in which a retailer that had cut prices had its supply terminated. The retailer claimed that termination occurred because of the price-cutting and that the termination constituted a violation of the antitrust laws. Although the Court stated that vertical agreements on resale prices are illegal per se, it ruled that because there was no agreement on price among the other competing retailers and the manufacturer, there was no violation of

⁸⁰*Dr. Miles Medical Company v. John D. Park & Sons Company*, 220 U.S. 373 (1911).

⁸¹The setting of maximum prices is judged under the rule of reason. See *State Oil v. Kahn*, 522 U.S. 3 (1997).

⁸²*Business Electronics Corporation v. Sharp Electronics Corporation*, 485 U.S. 717 (1988).

the antitrust laws. Therefore, although apparently the Court did not overrule its per se prohibition on vertical price fixing, its decision in this case reaches the conclusion of many economists that a manufacturer's control of pricing should not necessarily be an antitrust violation.

Exclusive Territories: A manufacturer may find it profitable to assign a geographic area, an *exclusive territory*, to one of its dealers and not allow its other dealers to locate in that area (Chapter 12). Exclusive territories provide dealers with incentives to promote the product and prevent one dealer from free riding on the promotional efforts of another. Exclusive territories can also adversely affect competition if they facilitate a cartel. Obviously, a territorial restriction on the ability of a manufacturer's dealers to compete literally restricts competition, even though the purpose of the territorial restriction may be to promote competition and the sale of the product.

In 1963, the Court addressed the issue of territorial restrictions in *White Motor*.⁸³ A truck manufacturer limited the territory in which its distributors could sell the product. The Court ruled that such territorial restrictions do not necessarily violate the antitrust laws and their legality should be determined only after examining their effects.

In *General Motors*, the Court investigated the location clauses that General Motors had in its dealers' contracts that prevented dealers from moving from one territory to another.⁸⁴ General Motors also tried to prevent its dealers from reselling cars to discount dealers, who sold them without the same promotional activities as other dealers. The Court ruled that the efforts of General Motors "to eliminate sales of new Chevrolet cars by discounters was to protect franchise dealers from real or apparent price competition." Accordingly, the Court ruled that this behavior violated the antitrust laws.

In *Schwinn*, the Court ruled that exclusive territories "are so obviously destructive of competition that their mere existence is enough."⁸⁵ This important case made the use of exclusive territories a per se violation of the antitrust laws.

In *Sealy* and *Topco*, the Supreme Court interpreted territorial restrictions as agreements to limit competition among rivals.⁸⁶ In both cases, groups of firms combined and agreed to territorial restrictions as part of an effort to promote their products and a common trademark and to avoid free-rider problems. The Court held that in both cases, these agreements were per se violations of Section 1 of the Sherman Act. However, to the extent that in both cases the territorial restrictions were necessary to develop a new trademarked product, it would seem that the subsequent ruling in the *BMI* case (agreements among firms are acceptable if they are necessary to provide the product) would mean that the *Sealy* and *Topco* cases would, if examined now, be analyzed under the rule of reason and not viewed as per se violations.

⁸³*White Motor Company v. United States*, 372 U.S. 253 (1963).

⁸⁴*United States v. General Motors Corp.*, 384 U.S. 127 (1966).

⁸⁵*United States v. Arnold, Schwinn & Company*, 388 U.S. 365 (1967).

⁸⁶*United States v. Sealy, Inc.*, 388 U.S. 350 (1967) and *United States v. Topco Associates Inc.*, 405 U.S. 596 (1972).

In 1977, the Court overruled *Schwinn* in *GTE Sylvania*.⁸⁷ Sylvania imposed locational restrictions on its distributors. The Court recognized that vertical restrictions improved the ability of a manufacturer to sell its product and provided a way to overcome certain free-rider problems (Chapter 12). Therefore, the Court overruled *Schwinn*'s per se prohibition against territorial restrictions and instead instituted a rule of reason under which vertical restrictions should be judged.

In *GTE Sylvania*, the Court's reasoning was based on the promotion of interbrand competition (competition among different products) at the expense of restricting intrabrand competition (competition among dealers of the same product). The use of this distinction is misleading. Vertical restrictions can indeed promote interbrand competition by making it profitable for dealers to promote and service each product, but it is not obvious that there is an undesirable effect on intrabrand competition. Although it is true in the literal sense that exclusive territories restrict the ability of one distributor to compete with another distributor, it is not true that a single manufacturer uses exclusive territories to restrict competition solely to raise the retail price and inflict an anticompetitive injury on consumers (Chapter 12).

After all, a manufacturer can raise the retail price (assuming no constraint from other products) by raising the wholesale price even without vertical restrictions. Through its control of the wholesale price, the manufacturer affects the retail price everywhere its product is sold. Rather than allowing the manufacturer to control only price, vertical restrictions give the manufacturer more control over promotional activities and service. By instituting a rule of reason criterion in *GTE Sylvania*, the Court acknowledged that vertical restraints can promote competition. See Example 19.6.

Exclusive Dealing: The Court has also analyzed *exclusive dealing* in which a manufacturer prevents its distributors from selling competing brands. Exclusive dealing allows manufacturers to overcome a different type of free-riding problem than the one overcome through the use of exclusive territories (Chapter 12). Exclusive territories address free riding of one dealer on the efforts of another; exclusive dealing addresses free riding of one manufacturer on the efforts of another. Exclusive dealing can also be used to raise entry barriers of rivals by raising distribution costs.

In 1922, the Supreme Court refused to enforce a manufacturer's contract with a retailer that forbade the retailer to sell brands of other manufacturers.⁸⁸ In 1949, in *Standard Stations*, the Court again addressed the problem of exclusive dealing.⁸⁹ Standard Oil of California required its independent dealers to purchase petroleum products and automobile accessories only from it. Rather than applying a rule of reason, the Supreme Court concluded that it would be too great a burden to show that competition had actually been diminished by the exclusive dealing and therefore it ruled that "Section 3 [of the Clayton Act, which forbids exclusive dealing] is satisfied by proof that competition has been foreclosed and a substantial share of the line of com-

⁸⁷*Continental TV Inc. v. GTE Sylvania Inc.*, 433 U.S. 36 (1977). See Preston (1994).

⁸⁸*Standard Fashion Company v. Magrain-Houston Co.*, 258 U.S. 346 (1922).

⁸⁹*Standard Oil Company of California v. United States*, 337 U.S. 293 (1949).

EXAMPLE 19.6*The FTC Plays with Toys ‘R Us*

Toys ‘R Us told toy manufacturers that if an *identical* toy were sold to warehouse clubs (very large stores that sell, often in bulk, at low prices), Toys ‘R Us would consider not carrying those particular toys. The FTC charged that this policy was designed to eliminate rival toy retailers in an effort to restrict competition.

Toys ‘R Us responded by claiming that they had no market power in toy retailing and that their policy was designed to limit free riding on their promotional activities by warehouse stores. It is difficult to predict which toys will be the Christmas season’s big hits. Typically, toy manufacturers do not raise the price of the hits. Instead, the manufacturers ration them to retail toy stores. An allocation of hits can be viewed as a payment by the manufacturers to toy stores that engage in extensive promotional activities, provide showroom services, and sell toys year round.

Toys ‘R Us engages in promotion and collaborates with manufacturers in designing toys. It stocks several thousand individual toy items in each store throughout the year, even though over 60 percent of toy sales occur in the last quarter of the year, the Christmas season. In contrast, warehouse clubs engage in little or no promotional activities and typically stock 100 to 150 toy items only at the end of the year.

Toys ‘R Us had about a 20 percent share of retail toy sales, while warehouse clubs had less than 5 percent. According to statistical evidence presented at trial, Toys ‘R Us did not have a statistically significant ability to raise retail prices even in areas where it faced only one major rival.

Despite this evidence, Toys ‘R Us lost the case. Its market share of toys has continued to fall and was about 17.5 percent in 2003.

Note: Carlton served as an expert witness for Toys ‘R Us.

Source: Carlton and Sider (1999); George James, “For Toys ‘R Us, A Time to Rebuild,” *New York Times*, January 14, 2004:1.

merce affected.” Justice Jackson dissented, arguing that the Court had made an error in economic reasoning and that exclusive dealing can be “a device for waging competition.”⁹⁰ The Court’s future treatment of exclusive dealing will presumably incorporate its *GTE Sylvania* decision, in which it recognized that vertical restrictions can promote competition sometimes.

Litigating Vertical Restraint Cases. In many cases, a distributor complains that a vertical agreement between some distributors and a manufacturer is intended to eliminate or prevent competition by other distributors. Plaintiffs phrase the complaint in

⁹⁰See also *Tampa Electric Company v. Nashville Coal Company*, 365 U.S. 320 (1961).

this way so as to characterize the behavior as a conspiracy involving price-fixing or output restrictions. The reason for this is that plaintiffs hope to apply the per se rules of the antitrust laws against price and output agreements and to obtain treble damages. If only breach of contract or other contract law violations were alleged, for example, they would receive only single damages. Thus, a terminated dealer may claim that the termination violates the antitrust laws rather than contract law in order to recover greater damages.

One example of an attempt to turn a vertical restriction case into a conspiracy case is *Klor's*, concerning the inability of a Klor's appliance store to obtain supplies from the same sources as its competitors.⁹¹ Klor's was located close to a Broadway Hale department store, which distributed appliances. Many well-known brands of appliances were sold to Broadway Hale, but not to Klor's. Klor's claimed that there was a conspiracy among Broadway Hale and the appliance manufacturers to drive it out of business. Klor's claimed that Broadway Hale used its market power to prevent manufacturers from selling to Klor's. Broadway Hale's defense indicated that numerous other retailers located close to Broadway Hale also sold the appliances of major manufacturers. The important economic question is whether Broadway Hale had sufficient buying power over the manufacturers to prevent them from selling to Klor's. If it did not, then a likely alternative explanation for the manufacturers' behavior is to control free-rider problems (see Chapter 12).

Price Discrimination

Many forms of price discrimination have been challenged under the antitrust laws. For example, predation can involve a firm's charging a lower price in a market where it faces a rival than in another market where it does not. Such price discrimination that harms direct competitors is called *primary-line price discrimination* (see for example, *Utah Pie*). A second form of price discrimination, *secondary-line price discrimination*, is one that leads to harm among the customers. The Robinson-Patman Act forbids both types of price discrimination. A third form of price discrimination that is restricted under antitrust laws is tie-in sales. This section discusses secondary-line discrimination and tie-in sales.

Price Discrimination Under Robinson-Patman

The Robinson-Patman Act prohibits a firm from price discriminating if it harms competition among the firm's customers (secondary-line discrimination). The Robinson-Patman Act (which amended Section 2 of the Clayton Act in 1936) was passed in response to political pressure from small retail stores (for example, grocery stores) that complained that larger chains were able to purchase supplies on more favorable terms

⁹¹*Klor's, Inc. v. Broadway Hale Stores Inc.*, 359 U.S. 207 (1959).

and thereby charge lower prices (Ross 1984). Many economists view the Robinson-Patman Act as special-interest legislation designed to protect small firms from competition from larger, more efficient firms that would be able to purchase supplies at low cost in the absence of the Act (Posner 1976b, Ross 1984).

One consequence of the Robinson-Patman Act is higher prices to consumers, who are deprived of the benefits of economies of scale in purchasing that the chain stores would otherwise be forced by competition (among themselves) to pass along to consumers (Ross 1984). The Robinson-Patman Act has led to substantial litigation (although government litigation has waned recently) and has also distorted pricing in many markets (Elzinga and Hogarty 1978). This law has harmed consumers. Although the FTC has brought relatively few cases in recent years, private actions are still brought.

Tie-in Sales

The antitrust laws have been used to prevent a firm from using *tie-in sales* in which the sale of one product is conditioned upon the purchase of another. The courts often characterize tie-in sales as a way of denying competitors the opportunity to make sales. Tie-in sales can arise for efficiency reasons or because a firm has some market power in one market and by the use of tie-in sales is able to earn higher profits than if it could only charge for one product (Chapter 10). Tie-in sales, then, can be a variant of price discrimination. They raise the return to being a monopoly in an industry. They do not necessarily create greater inefficiency losses and could result in output expansion. However, as discussed in Chapter 11, tie-in sales can also be used strategically as a tool to harm rivals. Curiously, courts have focused on cases more closely related to price discrimination.

Aside from the price discrimination motive, many products are naturally and efficiently tied together or bundled. For example, a car consists of many component parts, as does a radio. Consumers would be extremely unhappy if the government prohibited such efficient bundling of components. There are, however, some cases where tie-ins are anticompetitive. For example, tie-in sales could be used to raise entry costs (by tying repair to a machine, no independent repair shops arise and entrants are disadvantaged). However, the courts do not focus solely on these types of cases in their decisions to prohibit tie-in sales.

In early cases involving patented products, the Supreme Court ruled that tie-in sales were indeed legal. For example, in *A. B. Dick*, the Court did not find fault with A. B. Dick's practice of selling its patented mimeograph machines with a requirement that only ink purchased from A. B. Dick Company, as well as other supplies made by A. B. Dick, could be used.⁹² Such a tie-in could enable A. B. Dick to identify and extract more money from those who used the machine most intensively (Chapter 10).

The Clayton Act, passed in 1914, contained a section outlawing tie-in sales that had the effect of reducing competition. Soon thereafter, the Court overruled the *A. B.*

⁹²*Henry v. A. B. Dick Company*, 224 U.S. 1 (1912).

Dick case in *Motion Picture Patents*.⁹³ In two subsequent cases, the Supreme Court refused to alter its prohibition against tie-ins. In another, *IBM*, the United States attacked IBM's practice of selling key-punch machines with the requirement that the purchaser use only IBM tabulating cards.⁹⁴ IBM granted a special exception to the government that allowed it to use tabulating cards of its own manufacture provided the government paid an extra 15 percent rent. The Supreme Court used the government's experience to reject IBM's claim that its reputation would be damaged unless its tabulating cards were used because otherwise its machines might malfunction.

In *International Salt*, the Court investigated the requirement by International Salt Company that its purchasers use salt provided by International Salt in some machines that International Salt provided.⁹⁵ As in *IBM*, the Court rejected International Salt's claim that its reputation would be damaged if low-quality salt were used, with resulting damage to its machines. The Court ruled that since a substantial amount of the salt market was "foreclosed" to competitors, the tie-in was per se illegal.

In *Northern Pacific*, the Court ruled that the Northern Pacific Railway Company's requirements that lessees of certain lands be required to use its railway to ship under certain conditions was a per se violation.⁹⁶

Tying agreements serve hardly any purpose beyond the suppression of competition.

They deny competitors free access to the market for the tied product, not because the party imposing the tying requirement has a better product or a lower price, but because it has power leverage in another market.

The Court found that the defendant possessed substantial economic power and concluded that the tie-in was illegal. In *Fortner II*, the Court stated that "for a tie-in to be illegal, the seller must have some advantage not shared by his competitors in the market for the tie-in product."⁹⁷

Another important case involving tie-ins is *Hyde*.⁹⁸ A hospital had contracted for the provision of anesthesiology services from a private firm. The hospital agreed to use only that firm in the provision of anesthesiology services to its patients. An anesthesiologist sued and charged the hospital with tying anesthesiology to its other hospital services. The Court states that the requirements for an illegal tie are (1) the existence of two products, (2) market power in one product, and (3) *forcing*. By forcing, the Court means that products get sold together that would not be sold together without the tie. The Court said, "It is far too late in the history of our antitrust jurisprudence to question the proposition that certain tying arrangements pose an unacceptable risk of stifling competition. . . ." The Court did not find that

⁹³*Motion Picture Patents v. Universal Film Manufacturing Co.*, 243 U.S. 502 (1917).

⁹⁴*IBM Corporation v. United States*, 298 U.S. 131 (1936).

⁹⁵*International Salt Company v. United States*, 332 U.S. 392 (1947). See Peterman (1979) for an economic analysis of this case.

⁹⁶*Northern Pacific Railway Company v. United States*, 356 U.S. 1 (1958).

⁹⁷*United States Steel Corporation v. Fortner Enterprises, Inc.*, 429 U.S. 610 (1977).

⁹⁸*Jefferson Parish Hospital District No. 2 v. Hyde*, 466 U.S. 2 (1984). Lynk (1994a) analyzes this case.

the questioned conduct was illegal, apparently because the hospital lacked market power.

Kodak is an important tie-in case.⁹⁹ Kodak sells photocopiers in competition with many other firms. Kodak also provided Kodak parts and service to its customers. Kodak refused to supply certain parts to independent repair shops and was charged with illegally tying the sale of its photocopiers with its parts and service. Kodak asked that the case be dismissed because both sides agreed that Kodak faced lots of competition in the initial sale of photocopiers. If there is competition initially, then customers will not buy from Kodak if they know that they will be overcharged on repair parts and service. The Court rejected Kodak's argument. According to the Court, even if Kodak lacked market power initially in photocopiers, it is theoretically possible that either consumers are uninformed or that they are unable to forecast their repair cost. Essentially, the Court ruled that any equipment manufacturer could be considered a monopolist of its own unique repair parts and that a factual investigation is necessary to resolve a tie-in case even if there are hundreds of competing manufacturers of equipment. The Court failed to explain how any consumer would benefit if Kodak were forced to sell repair parts to independent repair shops at a price that Kodak could choose.

The courts' rulings regarding tie-in sales, especially those between franchisors and franchisees, have been inconsistent with the economic theory described in Chapter 12. A franchisor may place many types of restrictions on a franchisee in order to obtain certain desired economic performance. The franchisor also needs to be compensated for its efforts. One method by which a franchisee can compensate the franchisor is through various tie-in sales. For example, the franchisor could require that it be the sole supplier of napkins to the franchisee in lieu of charging a franchise fee based on sales. This conduct, however, has been attacked as a means to foreclose competition in napkins, using antitrust laws that ban tie-in sales.¹⁰⁰

Effects of Antitrust Laws on the Organization of Unregulated and Regulated Firms

When laws prohibit firms from taking particular actions, firms seek alternate routes to accomplish their objectives. Because the antitrust laws inhibit agreements between independent firms both horizontally and vertically, some firms merge or simply grow large and do everything themselves. Thus, the antitrust laws can encourage firms to

⁹⁹*Eastman Kodak Co. v. Image Technical Services, Inc.*, 112 S. Ct. 2072 (1992). See Calkins (1993) and Carlton (2001) for analyses of this case. See also the discussion of the Microsoft case, Example 11.5.

¹⁰⁰See, for example, *Siegel v. Chicken Delight Inc.*, 448 F.2d (9th Cir. 1971) cert. denied, 405 U.S. 955.

merge or grow larger than would otherwise occur. For example, Bittlingmayer (1985) explains that many firms merged around the turn of the century when antitrust laws first forbade agreements among firms but did not forbid mergers.

Similarly, some of the decisions of the Supreme Court to forbid certain types of contractual vertical restrictions create an incentive for vertical integration.¹⁰¹ Several states have responded by adopting statutes that prevent certain manufacturers from integrating forward into distribution. For example, several states prevent oil companies from owning and operating their own gasoline stations.

The antitrust laws, as already mentioned, affect how a franchisor can deal with a franchisee. The effect of the antitrust laws (and several state franchise laws) is to transfer certain rights from franchisors to franchisees and to make the franchise arrangement less attractive as a method of distribution (Smith 1982). If laws make it difficult for franchisors to control franchisees' actions, the incentive to use this mode of organization is diminished. Where antitrust laws encourage firms to alter their organizational form, it is likely that the new form will be less efficient.

Antitrust laws can have a large impact on regulated firms. The intersection of antitrust law and regulation is a complicated and contentious subject. Courts do not accept the argument that regulation immunizes a firm from antitrust actions because it is under the watchful eye of a regulator. Indeed, it was application of the antitrust laws that eventually forced the break-up of the AT&T phone monopoly in the early 1980s.¹⁰²

However, courts do accept the principle that regulation can immunize some actions of a regulated firm, especially when those actions are essential to the regulatory purpose. See for example, *Silver v. New York Stock Exchange*, 373 U.S. 341. In a case in 2003 (*in re: Stock Exchanges Options Trading Antitrust Litigation*, 317 F.3d 134), the Court ruled that the antitrust laws could not be used to attack the exchanges where options are traded for certain behavior because their actions were closely monitored by the Securities and Exchange Commission, the regulatory body charged with overseeing the operations of U.S. financial markets.

If a regulator imposes on a regulated firm a duty to deal with its rivals, then the rivals will be more formidable competitors. If a regulated firm misbehaves toward its rivals, should it be subject to antitrust sanctions or only to sanctions imposed by the regulator? The Supreme Court has recently ruled that the antitrust laws do not apply in cases where a firm is forced by regulation to deal with its rivals when it would not have done so in the absence of regulation.¹⁰³

¹⁰¹Alternatively, firms could use devices whose legality turns on legal technicalities. For example, a distributor may sell a good on consignment (which means that the manufacturer, not the distributor, owns the good) rather than owning the good and reselling it. The restrictions that can be placed on consignment sales can differ from those on nonconsignment sales.

¹⁰²*U.S. v. AT&T Co.*, 552 F. Supp. 131 (1982). See also www.aw-bc.com/carlton_perloff "The Breakup of AT&T."

¹⁰³See *Goldwasser v. Ameritech Corp.* 222 F.3d 390 (Seventh Circuit 2000) and *Verizon Communications Inc. v. Law Offices of Curtis V. Trinko*, 540 U.S. (2004).

SUMMARY

The major federal antitrust statutes are the Sherman Act, the Clayton Act, and the Federal Trade Commission Act. The interpretation of these statutes has varied considerably over time. There has been an increasing emphasis on the use of economic analysis in deciding what the antitrust laws should prohibit. Economists stress using antitrust laws to achieve efficiency.

It is a common mistake to think that the antitrust laws prohibit monopoly. They do not; however, they do prohibit certain actions that could allow a firm to acquire or maintain monopoly power.

Many antitrust cases revolve around whether a firm has market power, which is the ability to set price profitably above the competitive price. It is often difficult to assess directly whether a firm has market power. Courts and economists often use market share as a rough guide to whether a firm has market power. For this calculation of market share to be meaningful, the market must be properly defined. The market definition should include all those products whose presence significantly constrains the price of the product under analysis.

The courts use both per se rules and rules of reason. A per se rule prohibits certain acts without regard to the effect of the acts. For example, a price-fixing conspiracy whose sole purpose is to raise price is a per se violation. A rule of reason requires an investigation of the effect of the challenged conduct. Vertical restraints, other than on price, are now judged under a rule of reason.

The antitrust laws severely limit the types of cooperative behavior in which competitors can engage. For example, any attempt to fix price or limit output so as to harm consumers is a per se violation. That is, even unsuccessful attempts to fix price violate the law. There are some instances, however, where the courts allow cooperative behavior, even with respect to price, if the cooperative behavior is essential to producing the product. Mergers among competitors can be prevented if the effect of the merger is to create additional market power. Curiously, two firms with no market power are allowed to merge even though those same two firms would violate the law if they remained independent but spoke to each other and set price together.

The antitrust laws also constrain actions designed to hamper a firm's rivals. For example, strategic behavior, such as predatory pricing, designed to drive a rival out of business is illegal. The problem with antitrust enforcement in this area is that it is difficult to distinguish vigorous competition from strategic behavior that harms consumers. Overzealous enforcement could deprive consumers of the benefits of competition.

An important application of the antitrust laws has been to vertical relations between firms. The Court's reasoning has often been confused and has relied on a foreclosure of competition doctrine in which one firm that, say, vertically integrates into steel production is said to foreclose other steel producers from selling steel to the first firm.

There is a variety of reasons why a firm vertically integrates or imposes vertical restraints on its distributors. Many, but not all, of these reasons promote competition. Recently, the Court has recognized the possible procompetitive effect of nonprice vertical restrictions, but still regards vertical restrictions on price as a per se violation.

It is impossible to prove that vertical restrictions always benefit each consumer. The welfare effect of some vertical restrictions is ambiguous. Even after careful study, an analyst may be unable to decide whether a particular vertical restriction harms consumers. A vertical restriction may help some consumers and harm others. But the same could be said of the choice of product quality. There is often no greater justification for controlling how a manufacturer distributes its product than there is in dictating the quality of product that the monopoly produces.

There are, however, cases where vertical integration or restrictions harm consumers. Where the vertical integration or restrictions significantly impede or foreclose entry by rivals or where they allow distributors or manufacturers to act like a cartel, they harm consumers.

The welfare effects of price discrimination and tie-in sales designed to achieve price discrimination are generally ambiguous. As in the case of certain vertical restrictions, it is often costly and difficult to determine conclusively whether consumers are harmed in a particular situation. Pursuing strenuous general antitrust enforcement in areas with ambiguous welfare effects is unwise.

Using the antitrust laws to control some activities but not others can lead firms to adopt inefficient organizational forms. For example, if antitrust laws do not allow certain vertical restrictions but do allow vertical integration, firms may choose to vertically integrate to achieve their goals even if vertical integration is more costly than relying on vertical restrictions. Application of the antitrust laws to regulated industries can have large impacts on its market structure.

PROBLEMS

1. In *United Shoe*, the Court ruled that the 10-year leases adversely affected competition, in part, because they prevented other competitors from selling to a customer who had a United Shoe lease. Such a claim can be made of any contract. Identify circumstances where a long-term lease impedes competition.
2. Suppose Firm A is the only one that can sell in New York. Firm A faces competition elsewhere in the country. If Firm A can price discriminate, will the prices in New York differ from those elsewhere? Will there be a high correlation of price movements between New York and elsewhere? Is there one geographic market or two?
3. Suppose there are some industries in which the competitive equilibrium does not exist (the core does not exist—see www.aw-bc.com/carlton_perloff “Theory of the Core”). Should firms in

these industries be allowed to collude under the antitrust laws?

4. Where demand curves are compensated (adjusted for income effects), it can be shown that

$$\frac{\partial Q_j}{\partial p_i} = \frac{\partial Q_i}{\partial p_j},$$

where Q is the quantity demanded, p is the price, and subscripts indicate the products i or j . Suppose $p_i = p_j$, but $Q_i = 100Q_j$. What are the relative sizes of the two relevant cross-elasticities of demand? Why does it matter which one is used in the analysis of market definition?

5. It can be shown that

$$\frac{p_i}{Q_i} \frac{\partial Q_i}{\partial p_i} = - \sum_{j \neq i} \frac{p_j}{Q_i} \frac{\partial Q_i}{\partial p_j},$$

where Σ is a summation sign (sum over all products j other than product i), p is price, and Q is the compensated demand. Explain how this relation can be used to relate the elasticity of demand to cross-elasticities of demand. Use the formula to determine which cross-elasticity an analyst investi-

gating market power in Product A should examine to determine whether Product B constrains the pricing of Product A.

Answers to the odd-numbered problems are given at the back of the book.

SUGGESTED READINGS

Kwoka and White (2004) is an easy-to-read collection of articles providing economic analysis of some recent antitrust cases. Interesting books on the economics of antitrust law include Posner (2001), Posner and Easterbrook (1980) with subsequent supplements, and Williamson (1987). The Areeda and Hovenkamp (1997) treatise on antitrust pro-

vides an exhaustive analysis of antitrust issues. Pittman (1992) discusses merger law in Central and Eastern Europe, as do numerous Web sites sponsored by the International Competition Network. Carlton (2004b) discusses the lessons foreign countries have learned from U.S. antitrust experience.